Beyond Intervals: Phase Transitions Lead to More General Ranges

Karen Villaverde
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003
Email: kvillave@cs.nmsu.edu

Gilbert Ornelas
Computing Sciences Corporation
El Paso, TX 79912
Email: gtornelas@gmail.com

1. Objectives of Science and Engineering

- One of the main tasks of science and engineering:
  - use the current values of the physical quantities $x_1, \ldots, x_n$
  - to predict the future values y of the desired quantities.
- To be able to perform this prediction, we must know how y depends on x_i: $y = f(x_1, \ldots, x_n)$.
- Once we know the algorithm f and the values $x_1, \ldots, x_n$, we can predict y as $y = f(x_1, \ldots, x_n)$.
- *Comment.*
  - in this paper, we assume that we know the exact dependence f.
  - In reality, often, the algorithm f represents the actual physical dependence only approximately.

2. Measurement Inaccuracy

- In practice, the values of the quantities $x_1, \ldots, x_n$ usually come from measurements.
- Measurements are never 100% accurate.
- The measured value $\widetilde{x}_i$ is different from the (unknown) actual value x_i: $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0$.
- Usually, the manufacturer of the measuring instrument provides an upper bound Δ_i on Δx_i: $|\Delta x_i| \leq \Delta_i$.
- In this case, from the measurement result $\widetilde{x}_i$, we conclude that x_i is in the interval $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- *Conclusion:* we usually know the current values of the physical quantities with interval uncertainty.
- *Comment:* sometimes, we also know the probabilities of different values $x_i \in \mathbf{x}_i$.

3. The Effect of Measurement Inaccuracy on Prediction

- *Idealized case:*
  - we know the exact values x_i of the current quantities.
  - we can compute the exact value $y = f(x_1, \ldots, x_n)$ of the desired future quantity.
- *In practice:*
  - for each i, we only know the interval $\mathbf{x}_i$ of possible values of x_i;
  - then, we can only conclude that y belongs to the set

$$\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=}$$
$$\{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

4. Fuzzy Uncertainty: From the Computational Viewpoint, It Can Be Reduced to the Crisp Case

- For each possible value of $x_i \in \mathbf{x}_i$, the experts describe the degree $\mu_i(x_i)$ to which this value is possible.
- *Objective:* compute the fuzzy number corresponding to the desired value $y = f(x_1, \ldots, x_n)$.
- *Fact:* a fuzzy set can be thus viewed as a nested family of its α-cuts $\mathbf{x}_i(\alpha) \stackrel{\text{def}}{=} \{x : \mu_i(x) \geq \alpha\}$.
- *Meaning:* α-cut is the set of values of x_i which are possible with degree $\geq \alpha$.
- *Fact:* for each α, to α-cut $\mathbf{y}(\alpha)$ can be obtained by the interval formula: $\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_n(\alpha))$.
- *Conclusion:* to find the fuzzy number for y, we can apply an interval algorithm to the α-cuts $\mathbf{x}_i(\alpha)$.

5. Traditional Assumption: All Physical Dependencies are Continuous

- *Traditionally:* it is assumed that all the processes are continuous.
- *In particular:* that the function $y = f(x_1, \ldots, x_n)$ computed by the algorithm f is continuous.
- *Known:* the range of a continuous function on a bounded connected set, e.g., on $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$, is an interval.
- *Thus:* for continuous functions f, the range $\mathbf{y}$ of possible values of the future quantity y is an interval.
- *So:* due to inevitable measurement inaccuracy, we can only make predictions with interval uncertainty.
- Computing such intervals is one of the main tasks of *interval computations*.

6. Discontinuous Dependency: Physical Possibility

- *Example* of a discontinuous physical process: phase transition.
- When a water is heated and boils, its density abruptly decreases to the density of steam ρ_s.
- Of course, all the molecules move continuously.
- Theoretically, we continuously change from the density of water ρ_w to the density of steam ρ_s.
- However, for all practical purposes, this transition is very fast, practically instantaneous.
- So, in practice, we can safely assume that:
  - the future density ρ can be equal to ρ_w,
  - the future density ρ can be equal to ρ_s, but
  - the future density ρ cannot be equal to any intermediate value.

7. How Discontinuities Affect the Class of Possible Ranges?

- *First idea:* it is sufficient to consider closed ranges.
- *Motivation:*
  - let $s_1, s_2, \ldots, s_k, \ldots$ are all possible, and $s_k \to s$,
  - then for every accuracy there is s_k that is indistinguishable from s (and possible);
  - thus, from the practical viewpoint, s is also possible.
- *Second idea:* it is sufficient to consider closed classes of sets.
- *Motivation:* similar; as a measure of closeness between sets, we use Hausdorff metric:

$$d_H(S, S') = \min\{\varepsilon > 0 : S \text{ is in the } \varepsilon\text{-neighborhood of } S' \text{ and } S' \text{ is in the } \varepsilon\text{-neighborhood of } S\}.$$

8. Definitions and the Main Result

Definition. *A class $\mathcal{S}$ of closed bounded non-empty subsets of the real line is called a* class of ranges *if it satisfies the following conditions:*

(i) the class $\mathcal{S}$ contains an interval;

(ii) the class $\mathcal{S}$ is closed under arbitrary continuous transformations, i.e., if $S \in \mathcal{S}$ and $f(x)$ is a continuous function, then $f(S) \in \mathcal{S}$;

(ii) there exist a value x_0 and a function $f_0(x)$ which is continuously increasing for $x < x_0$ and for $x > x_0$ and which has a "jump" at x_0 ($f_0(x_0-) < f_0(x_0+)$) such that the class $\mathcal{S}$ is closed under f_0, i.e., if $S \in \mathcal{S}$ then $\overline{f_0(S)} \in \mathcal{S}$; and

(iv) the class $\mathcal{S}$ is closed (under Hausdorff metric).

Theorem. *The class of ranges coincides with the class of all bounded closed sets.*

9. Computational Complexity of the Prediction Problem: Interval Uncertainty, Linear Functions

- *Starting point:* interval uncertainty, linear function

$$y = f(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_i \cdot x_i.$$

- *Approximate value:* $\widetilde{y} = a_0 + \sum\limits_{i=1}^{n} a_i \cdot \widetilde{x}_i$.

- *Approximation error* $\Delta y = \widetilde{y} - y$ is $\Delta y = \sum\limits_{i=1}^{n} a_i \cdot \Delta x_i$, where $\Delta x_i \in [-\Delta_i, \Delta_i]$.

- $\sum\limits_{i=1}^{n} a_i \cdot \Delta x_i \to \max$ iff $a_i \cdot \Delta x_i \to \max$ for all i.

- *Conclusion:* the largest possible value of the sum Δy is $\Delta = |a_1| \cdot \Delta_1 + \ldots + |a_n| \cdot \Delta_n$.

- *Computational complexity:* linear time (i.e., efficient).

10. Computational Complexity of the Prediction Problem: Interval Uncertainty, Quadratic Functions

- *Case:* interval uncertainty, quadratic functions

$$y = f(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_i \cdot x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \cdot x_i \cdot x_j.$$

- *Given:* interval inputs $x_i \in \mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- *Compute:* the range

$$\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

- *Result:* this problem is NP-hard.
- *What is NP-hard:* (if P≠NP, then)
  - no feasible (polynomial time) algorithm
  - can compute the exact endpoints of the range $\mathbf{y}$
  - for all possible intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

11. Computational Complexity of the Prediction Problem: General Uncertainty, Linear Functions

- *Reminder:* due to discontinuities, the range $\mathbf{x}_i$ of x_i is, in general, different from the interval $[\underline{x}_i, \overline{x}_i]$.
- *Result (reminder):* the range can be equal to an arbitrary bounded closed set

$$S_i \subseteq [\underline{x}_i, \overline{x}_i].$$

- *Example:* this range can be equal to the 2-point set

$$\mathbf{x}_i = \{\underline{x}_i, \overline{x}_i\}.$$

- *New result:* for 2-point inputs, the problem of computing the range is NP-hard even for linear functions.
- *Comment:* the proof is in our proceedings paper.

12. Conclusions

- One of the main tasks of science and engineering is:
  - to use the current values of the physical quantities
  - for predicting the future values of the desired quantities.
- Due to the measurement inaccuracy, we usually know the current values with interval uncertainty.
- Traditionally, it is assumed that all the processes are continuous.
- As a result, the range of possible values of the future quantities is also known with interval uncertainty.
- In many practical situations, the dependence of the future values on the current ones is discontinuous.
- *Example of discontinuity:* phase transitions.

13. Conclusions (cont-d)

- *Reminder:* dependencies can be discontinuous.
- *Objective:* compute the range $\mathbf{y}$ of possible values of the future quantity y.
- *Main result:* initial interval uncertainties can lead to arbitrary bounded closed range $\mathbf{y}$.
- *Corollary:* discontinuity may drastically increase the computational complexity of computing $\mathbf{y}$.
- *Example:* for linear functions, the complexity increases
  - from linear time
  - to NP-hard.

14. Proof of NP-Hardness

- *Main idea:* reduce partition problem – known to be NP-hard – to our problem:
  - given n positive integers $s_1, \ldots, s_n$,
  - check whether $\exists \varepsilon_i \in \{-1, 1\}$ s.t. $\sum\limits_{i=1}^{n} \varepsilon_i \cdot s_i = 0$.
- *Reduction:* reduce each particular case of this problem to the following particular case of our problem:
  - compute the range $\mathbf{y}$ of $f(x_1, \ldots, x_n) = \sum\limits_{i=1}^{n} s_i \cdot x_i$
  - when $x_i \in \{-1, 1\}$.
- *Fact:* 0 belongs to the range $\mathbf{y}$ iff the original instance of the partition problem has a solution.
- The reduction is proven, hence our problem is indeed NP-hard.