From Gauging Accuracy of Quantity Estimates to Gauging Accuracy and Resolution of Measuring Physical Fields: A Broad Prospective on Fuzzy Transforms

Irina Perfilieva[1] and Vladik Kreinovich[2]

[1]University of Ostrava
Inst. for Research and
Applications of Fuzzy Modeling
ul. 30. dubna 22
701 00 Ostrava 1, Czech Republic
Irina.Perfilieva@osu.cz
[2]Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
vladik@utep.edu

Need for data processing. In many real-life situations, we are interested in the value of a quantity which is difficult (or even impossible) to measure directly. For example, we may be interested:

- in the distance to a faraway star, or
- in the amount of water in an underground water layer.

Since we cannot measure the corresponding quantity y directly, we measure it *indirectly*. Specifically,

- we find easier-to-measure quantities $x_1, \ldots, x_n$ which are related to the desired quantity y by a known dependence $y = f(x_1, \ldots, x_n)$;
- we measure the values of the auxiliary quantities $x_1, \ldots, x_n$; and
- we use the results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of measuring the auxiliary quantity to compute the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for the desired quantity y.

For example, to find the distance y to a faraway star, we can use the following *parallax* method (perfected in modern times by Tycho Brahe):

- we measure the orientations x_1 and x_2 to this star at two different seasons,
- we measure the the distance x_3 between the spatial locations of the corresponding telescopes at these two seasons (i.e., in effect, the diameter of the earth orbit);
- then, reasonably simply trigonometric computations enable us to describe the desired distance y as a function of the easier-to-measure quantities x_1, x_2, and x_3.

In general, computations related to such indirect measurements form an important particular case of data processing.

Need to take uncertainty into account. Measurements are never absolutely accurate. As a result, the measurement results $\widetilde{x}_i$ are, in general, different from the actual (unknown) values x_i of the measured quantities. Because of this, the result $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ of data processing is, in general, different from the actual (unknown) value $y = f(x_1, \ldots, x_n)$.

Thus, in practical applications, we need to take this uncertainty into account.

Interval uncertainty. In practice, we often only know the upper bound Δ_i on the measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$: $|\Delta x_i| \le \Delta_i$. In this case, the only information that we have about the actual values x_i is that x_i belongs to the *interval* $\mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

Under such interval uncertainty, we need to find the range of possible values of y: $\mathbf{y} = \{f(x_1, \ldots, x_n) : x_i \in \mathbf{x}_i\}$. The problem of computing this range is known as *interval computations*; see, e.g., [1].

Need to measure physical fields. In practice, the situation is often more complex: the values that we measure can be:

- values $v(t)$ of a certain dynamic quantity v at a certain moment of time t
- or, more generally, the values $v(x,t)$ of a certain physical field v at a certain location x and at a certain moment of time t.

For example, in geophysics, we are interested in the values of the density at different locations and at different depth.

Need to take uncertainty into account when measuring physical fields. When we measure physical fields,

- not only we get the measured value $\widetilde{v} \approx v$ with some inaccuracy, but
- also the location x is not exactly known.

Moreover, the sensor picks up the "averaged" value of v at locations close to the approximately known location $\widetilde{x}$.

In other words,

- in addition to inaccuracy $\widetilde{v} \neq v$,
- we also have a finite *resolution* $\widetilde{x} \neq x$.

Estimating uncertainty related to measuring physical fields: challenging problems. How can we describe the relation between these measurement results and the actual field values? How can we process such generalized interval data?

The answers to these questions clearly depend on what we know about the resolution.

Case of fuzzy transforms. In some cases, we know how the measured values $\widetilde{v}_i$ are related to $v(x)$, i.e., we know the weights $w_i(x)$ in the dependence

$$\widetilde{v}_i \approx \int w_i(x) \cdot v(x)\, dx.$$

In this case, all our information about $v(x)$ is contained in the set of values $\widetilde{v}_i$.

This situation was first considered in [4, 5, 6]. The values $\widetilde{v}_i$ corresponding to the (unknown) functions $v(x)$ are known as the *fuzzy transform* of the filed $v(x)$. In the first part of our talk, we describe how we can use the knowledge of the fuzzy transform to estimate the accuracy and resolution of the corresponding measurements.

Case of interval uncertainty. In other cases – similarly to the above interval setting – we only know the upper bound Δ_x on the location error $\widetilde{x} - x$. In this case, every measured pair $(\widetilde{v}_i, \widetilde{x}_i)$ must be (Δ, Δ_x)-close to some value $(v(x), x)$.

This closeness can be naturally described in terms of an (asymmetric) Hausdorff distance $d(\widetilde{V}, V)$ between the set of measurement pairs $\widetilde{V}$ and the graph V of the field $v(x)$. In the second part of our talk, we describe how we estimate the accuracy and resolution of the corresponding measurements in this case. Our approach uses techniques originally described in [1].

Case of minimal knowledge about uncertainty. Yet another case is when we do not even know Δ_x. It happened, e.g., when we solve the seismic inverse problem to find the velocity distribution.

In this case, a natural heuristic idea is:

- to add a perturbation of size Δ_0 (e.g., sinusoidal) to the reconstructed field $\widetilde{v}(x)$,
- to simulate the new measurement results,
- to apply the same algorithm to the simulated results, and
- to reconstruct the new field $\widetilde{v}_{\text{new}}(x)$.

If the perturbations are not visible in $\widetilde{v}_{\text{new}}(x) - \widetilde{v}(x)$, this means that details of size Δ_0 cannot be reconstructed and so, the actual resolution is $\Delta_x > \Delta_0$. This approach was partially described in [2, 7].

In the third part of the talk, we elaborate on this approach, and derive the optimal shape of perturbations.

References

[1] R. Anguelov, S. Markov, and B. Sendov, "The set of Hausdorff continuous functions – the largest linear space of interval functions", *Reliable Computing*, 2006, Vol. 12, pp. 337–363.

[2] M. G. Averill, *A Lithospheric Investigation of the Southern Rio Grande Rift*, PhD Dissertation, University of Texas at El Paso, Department of Geological Sciences, 2007.

[3] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[4] I. Perfilieva, “Fuzzy transform: application to reef growth problem”, In: R. Demicco and G. J. Klir (Eds.), *Fuzzy Logic in Geology*, Academic Press, Amsterdam, 2003, pp. 275–300.

[5] I. Perfilieva, “Fuzzy transforms: theory and applications”, *Fuzzy Sets and Systems*, 2006, Vol. 157, pp. 993–1023.

[6] I. Perfilieva, V. Novák, and A. Dvorák, “Fuzzy transform in the analysis of data”, *International Journal of Approximate Reasoning*, 2008, Vol. 48, No. 1, pp. 36–46.

[7] P. Pinheiro da Silva, A. Velasco, M. Ceberio, C. Servin, M. Averill, and N. Del Rio, “Propagation and Provenance of Probabilistic and Interval Uncertainty in Cyberinfrastructure-Related Data Processing and Data Fusion”, In: R. L. Muhanna and R. L. Mullen (eds.), *Proceedings of the International Workshop on Reliable Engineering Computing REC'08*, Savannah, Georgia, February 20–22, 2008, pp. 199–234.