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

Vladik Kreinovich[*] *and Irina Perfilieva*[†]

Abstract: Fuzzy transform is a new type of function transforms that has been successfully used in different application. In this paper, we provide a broad prospective on fuzzy transform. Specifically, we show that fuzzy transform naturally appears when, in addition to *measurement* uncertainty, we also encounter another type of *localization* uncertainty: that the measured value may come not only from the desired location x, but also from the nearby locations.

Key words: *fuzzy transform, interval uncertainty, interval computations, Hausdorff metric*

Received: !!!
Revised and accepted: !!!

1. Need for data processing

Idea. 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:

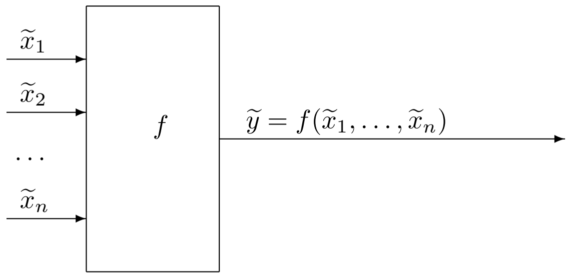
- in the distance to a faraway star, or
- in the amount of water in an underground water layer.

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

[*]Vladik Kreinovich
University of Texas at El Paso, El Paso, Texas, USA, `vladik@utep.edu`
[†]Irina Perfilieva
Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Czech Republic, `irina.perfilieva@osu.cz`

- 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.



Comment. In the simplest cases, we know an *explicit* analytical expression for the dependence $f(x_1, \ldots, x_n)$. In many other cases, we only have an *implicit* description of the dependence between the desired quantity y and the easier-to-measure quantities $x_1, \ldots, x_n$. For example, we may have a system of equations (or a system of differential equations) that relates y and x_i. In such situations, we usually have an algorithm for transforming the values $x_1, \ldots, x_n$ into the desired value y. For example, we may have have an algorithm that solves the corresponding system of differential equations.

In this paper, we consider the most general case of the dependence – when $f(x_1, \ldots, x_n)$ is an algorithm. The case of an explicit analytical dependence is also covered; in this case, we have a simple explicit algorithm for computing this expression.

Example. To find the distance y to a faraway star, we can use the following *parallax* method:

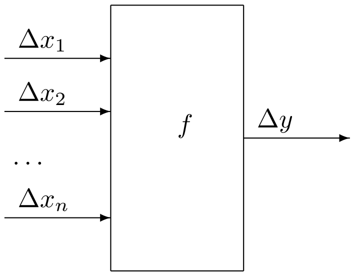
- we measure the orientations x_1 and x_2 to this star at two different seasons,
- we measure the the distance x_3 between the spatial locations of the corresponding telescopes at these two seasons (i.e., in effect, the diameter of the earth orbit);
- then, reasonably simply trigonometric computations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

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

2. 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: $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0$. 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 $y = f(x_1, \ldots, x_n)$: $\Delta y \stackrel{\rm def}{=} \widetilde{y} - y \neq 0$.



Thus, in practical applications, we need to take this uncertainty into account. !!! also f may be known with uncertainty

3. Probabilistic uncertainty and its limitations

Traditional probabilistic approach. Traditional approach to uncertainty estimation is based on the assumption that we know the probability distribution for the measurement errors Δx_i.

Usually, it is assumed that the measurement errors Δx_i are independently normally distributed, with zero means and known standard deviations σ_i; see, e.g., [10]. In this case, we can use standard statistical techniques to determine the probability distribution of the resulting measurement error Δy.

Probabilistic approach beyond normal distributions. In some cases, the distribution of the measurement errors is known to be non-Gaussian – and we know the exact shape of the corresponding probability distribution.

For example, if the main component of the measurement error comes from the sinusoidal electric filed generated by the electric plugs – then the measured errors are distributed according to the arcsine law.

In such cases, we can use statistical techniques – e.g., Monte-Carlo simulations (if no analytical techniques are known for this distribution) to find the desired probability distribution for Δy.

Where do probabilities come from? In some important practical situations, we do not know the probabilities of different values Δx_i. Indeed, in the measurement practice, these probabilities usually come from the *calibration* of the corresponding measuring instrument (MI), i.e., by comparing its measurement results with the results of measuring the same quantity by a much more accurate ("standard") measuring instrument.

Since the standard measuring instrument is much more accurate than the one that we are calibrating, its measurement error can be safely ignored in comparison with the measurement errors of our original MI. Thus, for each measurement k, the difference $\widetilde{x}^{(k)} - \widetilde{x}^{(k)}_{\rm st}$ between the values measured by the original and the standard

MI can serve as a reasonable approximation to the actual (unknown) measurement error $\widetilde{x}^{\,k} - x^{(k)}$. In this approximation, the statistics of the calibration differences $\widetilde{x}^{\,(k)} - \widetilde{x}^{(k)}_{\rm st}$ can serve as a good description of the probability distribution of the measurement error.

In many practical situations, this calibration is a standard measurement practice. In such situations, we do indeed know the probabilities of different values of Δx_i. However, there are important situations when this calibration is not done.

Situations when we do not know probabilities: fundamental science. The first such situation occurs in fundamental sciences, when we process state-of-the-art measurements. For example, the Hubble telescope provides unique state-of-the-art measurements of celestial bodies. It would be nice if there was a five time more accurate telescope floating nearby that we could use for calibration – but the Hubble telescope is the best we have.

Similarly, it would be nice to have a "standard" (more accurate) super-collider to calibrate the existing CERN colliders – but they are the best we have.

In such situations, we only know upper bounds on the measurement error. In such situations, we do not know the probabilities of different values Δx_i. What we do know is the *upper bound* Δ_i on the (absolute value of) the measurement error. Indeed, if we do not even know the upper bound, this means that difference the actual (unknown) value x_i and the measured value $\widetilde{x}_i$ can be as large as possible – so the value $\widetilde{x}_i$ is not a measurement, it is just a wild guess which can be completely wrong.

Another situation when we do not know probabilities: manufacturing. Another important practical situation when we do not know the probabilities of different values of Δx_i is the situation of manufacturing practice. In principle, we can calibrate every single sensor, every single measuring instrument. However, calibration is a very expensive process, involving expensive super-accurate "standard" measuring instruments. In manufacturing practice, where the profit margins are low, any unnecessary expense is avoided – in particular, most sensors are *not* calibrated. For such sensors, we do not know the probabilities of different values of measurement errors Δx_i, we only know the upper bounds Δ_i provided by the manufacturers of these sensors and measuring instrument.

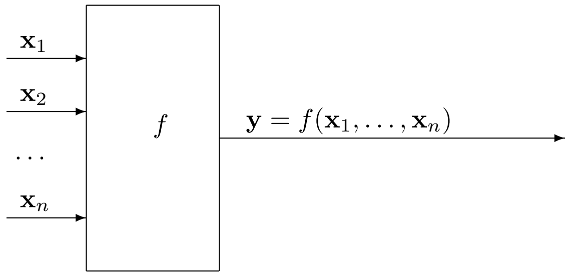
4. Interval uncertainty

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

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].



5. 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.

6. 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* $\widetilde{x} \neq x$.

7. Estimating uncertainty related to measuring physical fields: challenging problems

In general, the measured value $\widetilde{v}_i$ differs from the averaged value v_i by the measurement imprecision $\Delta v_i = \widetilde{v}_i - v_i$. In the interval case, we know the upper bound

Δ_i on this measurement error $|\Delta v_i| \le \Delta_i$. Thus, the averaged quantity v_i can take any value from the interval $[\underline{v}_i, \overline{v}_i]$, where $\underline{v}_i \stackrel{\text{def}}{=} \widetilde{v}_i - \Delta_i$ and $\overline{v}_i \stackrel{\text{def}}{=} \widetilde{v}_i + \Delta_i$.

Based on these bounds on v_i, what can we learn about the original field $v(x)$? The answer to this questions depends on what we know about the averaging, i.e., on the dependence of v_i on $v(x)$.

In principle, there are three possible situations:

- sometimes, we know exactly how the averaged values v_i are related to $v(x)$;
- sometimes, we only know the upper bound δ on the location error $\widetilde{x} - x$ (this is similar to the interval case);
- sometimes, we do not even know δ.

In the following sections, we describe how to process all these types of uncertainty.

8. Possibility of linearization

Sometimes, the signal $v(x)$ that we are measuring is large, i.e., the values of the signal are much larger than the noise (and the measurement errors in general). In such situations, the measured values well represent the actual signal, and for many applications, the measurement errors can be safely ignored.

The need to take into account measurement errors becomes important only when the signal $v(x)$ is relatively weak. In this case, we can expand the dependence of v_i on $v(x)$ in Taylor series and ignore quadratic and higher order terms in this dependence. As a result, we get a linear expression for v_i in terms of $v(x)$:

$$v_i = \int w_i(x) \cdot v(x)\, dx.$$

9. Case of full information about the resolution

Description. In this section, we consider the case when we know the exact expression for this dependence, i.e., when we know the weights $w_i(x)$.

The notion of fuzzy transform. Intuitively, each "averaged" value v_i can be viewed as the value of the field $v(x)$ at a "fuzzy" point characterized by uncertainty $w_i(x)$. Because of this interpretation, the transformation from the original function $v(x)$ to the set of values $v_1, \ldots, v_n$ is also known as a *fuzzy transform*; see, e.g., [6, 7, 8].

What we want to predict. Based on the measurement results $\widetilde{v}_1, \ldots, \widetilde{v}_n$, we would like to reconstruct the field $v(x)$. From the pragmatic viewpoint, knowing the field means being able to predict the results of all other measurements of this field.

Each such measurement can be characterized by its own averaging function $w(x)$. Thus, predicting the result of the measurement means predicting the corresponding averaged value $y = \int w(x) \cdot v(x)\, dx$.

Of course, the space of functions is infinite-dimensional, which means that to uniquely reconstruct a function, we need to know infinitely many parameters. Thus, based on n numbers $\widetilde{v}_1, \ldots, \widetilde{v}_n$, we cannot uniquely reconstruct the function $v(x)$ – and thus, we cannot uniquely reconstruct the desired averaged value y. So, the problem is to find the *range* $[\underline{y}, \overline{y}]$ of this value y.

Prediction problem as a particular case of linear programming. The lower endpoint $\underline{y}$ is the smallest possible value of y, the upper endpoint $\overline{y}$ is the largest possible value of y. Thus, the problems of finding the desired endpoints $\underline{y}$ and $\overline{y}$ can be formulated in the following optimization form:

Minimize (maximize) $y = \int w(x) \cdot v(x)\, dx$
under the constraints

$$\underline{v}_i \le \int w_i(x) \cdot v(x)\, dx \le \overline{v}_i, \quad 1 \le i \le n.$$

In both problems, we optimize the value of a linear functional under linear constraints, so from the mathematical viewpoint, these problems are (infinite-dimensional) linear programming problems.

Without prior restrictions on the field $v(x)$, we cannot predict anything. In general, if we do not impose any conditions on the function $v(x)$, then both bounds are infinite – unless $w(x)$ is a linear combination of $w_i(x)$.

Indeed, it is known that every vector w which is orthogonal to all the vectors t, which are orthogonal to all the vectors $w_1, \ldots, w_n$, belongs to the linear space generated by the vectors $w_1, \ldots, w_n$ – i.e., is a linear combination of $w_1, \ldots, w_n$. Thus, if a vector w cannot be represented as a linear combination of the vectors $w_1, \ldots, w_n$, then there exists a vector t which is orthogonal to all w_i but not to w. With respect to the space of all the functions, this means that if $w(x)$ cannot be represented as a linear combination of the functions $w_i(x)$, then there exists a function $t(x)$ which is orthogonal to all $w_i(x)$ (in the sense that $\langle w_i, t\rangle \stackrel{\text{def}}{=} \int w_i(x) \cdot t(x)\, dx = 0$) but not to $w(x)$ ($\langle w, t\rangle \ne 0$).

For an arbitrary real number λ, instead of the actual field $v(x)$, we can now consider a new field $v_\lambda(x) \stackrel{\text{def}}{=} v(x) + \lambda \cdot t(x)$. For this new field $v_\lambda(x)$, the values of v_i are the same as for the original field $v(x)$ – and hence, satisfy the same inequalities. However, the new value y is equal to $y_\lambda = \langle w, v\rangle + \lambda \cdot \langle w, t\rangle$. Since $\langle w, t\rangle \ne 0$, for appropriate λ, we can get this value y_λ equal to any given real number. Thus, indeed, the smallest possible value of y is $\underline{y} = -\infty$, and the largest possible value of y is $\overline{y} = +\infty$.

Non-negative fields. In many practical problems, the field $v(x)$ can only have non-negative values $v(x) \ge 0$. For example, in geophysics, the density $v(x)$ cannot be negative.

Under this additional restrictions, we already have non-trivial bounds $\underline{y}$ and $\overline{y}$.

Dual linear programming techniques. For solving these problems, we can use the experience of imprecise probabilities [4, 11] where in similar linear programming problems, $v(x)$ is the non-negative probability density function (and the weights are, e.g., functions x^2 corresponding to moments). According to this experience, many efficient algorithms come from considering *dual* linear programming problems, i.e., by computing the range $[\underline{v}, \overline{v}]$, where

$$\underline{v} = \sup\left\{\sum y_i \cdot \underline{v}_i : \sum y_i \cdot w_i(x) \le w(x)\right\};$$

$$\overline{v} = \inf\left\{\sum y_i \cdot \overline{v}_i : w(x) \le \sum y_i \cdot w_i(x)\right\}.$$

Indeed, if $\sum y_i \cdot w_i(x) \le w(x)$, then, by multiplying both sides of this inequality by $v(x) \ge 0$ and integrating over x, we conclude that $\sum y_i \cdot v_i \le y$. Since we know that $v_i \ge \underline{v}_i$, we thus get a lower bound for y: $y \ge \sum y_i \cdot \underline{v}_i$. Thus, y is larger than the largest of these bounds, i.e., $y \ge \underline{v}$. So, we can conclude that $\underline{y} \ge \underline{v}$. Similarly, we can conclude that $\overline{y} \le \overline{v}$, i.e., that the dual linear programming interval $[\underline{v}, \overline{v}]$ is the enclosure for the desired range $[\underline{y}, \overline{y}]$.

Comments.

- For discrete linear programming problems, the dual interval is exactly equal to the original one.
- Our problems are easier than the imprecise probability ones, since the functions $w_i(x)$ are usually localized and thus, for each x, usually at most a few functions $w_i(x)$ differ from 0. This makes checking the sums easier.
- Checking the inequalities like $\sum y_i \cdot w_i(x) \le w(x)$ is even easier in a practically important case of piece-wise linear functions $w_i(x)$ and $w(x)$. In this case, it is sufficient to check this inequality at endpoints of linearity intervals – then, due to linearity, it will be automatically true for all internal points as well.

10. Situations in which we only know upper bounds

General idea. In other cases – similarly to the interval setting – we do not only know the upper bounds δ on the location error $\widetilde{x} - x$. A natural question is: when is a model $v(x)$ consistent with the given observation $(\widetilde{v}, \widetilde{x})$?

In this case, the measured value $\widetilde{v}$ is Δ-close to a convex combination of values $v(x)$ for x s.t. $\|x - \widetilde{x}\| \le \Delta x$. Thus, $\underline{v}_\delta(\widetilde{x}) - \Delta \le \widetilde{v} \le \overline{v}_\delta(\widetilde{x}) + \Delta$, where:

$$\underline{v}_\delta(\widetilde{x}) \stackrel{\text{def}}{=} \inf\{v(x) : \|x - \widetilde{x}\| \le \delta\}, \text{ and } \overline{v}_\delta(\widetilde{x}) \stackrel{\text{def}}{=} \sup\{v(x) : \|x - \widetilde{x}\| \le \delta\}.$$

Case of interval models. In real life, we rarely have an *exact* model $v(x)$. Usually, we have *bounds* on $v(x)$, i.e., an interval-valued model $\mathbf{v}(x) = [v^-(x), v^+(x)]$. An observation $(\widetilde{v}, \widetilde{x})$ consistent with this "interval-valued" model if these exists a model $v(x) \in \mathbf{v}(x)$ which is consistent with this observation.

Since the values $\underline{v}_\delta$ and $\overline{v}_\delta$ monotonically depend on $v(x)$, this consistency leads to

$$\underline{v}_\delta^-(\widetilde{x}) - \Delta \le \widetilde{v} \le \overline{v}_\delta^+(\widetilde{x}) + \Delta.$$

Relation to Hausdorff metric. In many practical problems, the field $v(x)$ continuously depends on x. For continuous functions, inf and sup on a bounded closed set $\{x : \|x - \widetilde{x}\| \le \delta\}$ are attained at some value. Thus, the above criterion for consistency between a model and observations can be simplified.

Namely, in this case, the set $\widetilde{m}$ of all measurement results $(\widetilde{v}, \widetilde{x})$ is consistent with the model $v(x)$ if and only if

$$\forall (\widetilde{v}, \widetilde{x}) \in \widetilde{m}\, \exists (v(x), x) \in v\, ((\widetilde{v}, \widetilde{x}) \text{ is } (\Delta, \delta)\text{-close to } (v(x), x)),$$

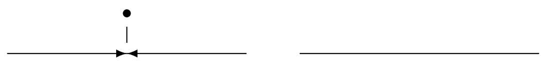
i.e., $|\widetilde{v} - v| \le \Delta$ and $\|x - \widetilde{x}\| \le \delta$.

This definition is similar to the standard definition of the *Hausdorff metric* d_H: $d_H(A, B) \le \varepsilon$ means that

$$\forall a \in A\, \exists b \in B\, (d(a, b) \le \varepsilon) \text{ and } \forall b \in B\, \exists a \in A\, (d(a, b) \le \varepsilon).$$

(This similarity was noticed in [1].)

Specifically, the above definition is an *asymmetric* version of Hausdorff metric. Let us show, on a simple example, that our "distance" is indeed asymmetric.

Case 1: 

In this example,

- the actual field has the form $v(0) = 1$ and $v(x) = 0$ for $x \ne 0$, and
- the measurements results are all zeros, i.e., $\widetilde{v} = 0$ for all $\widetilde{x}$.

In this case, all the measurements are consistent with the model:

- the values $\widetilde{v} = 0$ for $\widetilde{x} \ne 0$ are consistent with $v = 0$ for $x = \widetilde{x}$, and
- the value $\widetilde{v} = 0$ for $\widetilde{x} = 0$ is consistent with $v(x) = 0$ for $x = \delta$ s.t. $|\widetilde{x} - x| \le \delta$.

Case 2: 

In this example,

- the actual field is all zeros, i.e., $v(x) = 0$ for all x, and
- the measurement results are $\widetilde{v} = 1$ for $\widetilde{x} = 0$, and $\widetilde{v} = 0$ for all $\widetilde{x} \ne 0$.

Here, when $\Delta < 1$, the measurement $(1, 0)$ is *inconsistent* with the model, because for all x which are δ-close to $\widetilde{x} = 0$, we have $v(x) = 0$ hence we should have $|\widetilde{x} - v(x)| = |\widetilde{x}| \le \Delta$.

11. Case of minimal knowledge about uncertainty

Idea. Yet another case is when we do not even know δ. It happens, e.g., when we solve the seismic inverse problem to find the velocity distribution [2, 9].

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 $\widetilde{v}_{\rm new}(x)$.

If the perturbations are not visible in $\widetilde{v}_{\rm new}(x)-\widetilde{v}(x)$, this means that details of size δ_0 cannot be reconstructed and so, the actual resolution is $\delta > \delta_0$. This approach was partially described in [2, 9].

Linearization and its consequences. Which perturbations should we choose? To select the optimal perturbations, we will take into account the fact that since perturbations are usually small, we can safely linearize their effects. Thus, if we know the results $\Delta v_1(x)$, $\ldots$, $\Delta v_k(x)$ of applying perturbations $e_1(x)$, $\ldots$, $e_k(x)$, we can predict the result $\Delta v(x)$ of applying an linear combination

$$e(x) = c_1\cdot e_1(x)+\ldots+c_k\cdot e_k(x),$$

as

$$\Delta v(x) = c_1\cdot \Delta v_1(x)+\ldots+c_k\cdot \Delta v_k(x).$$

In other words, once we know the results of applying k different perturbations $e_1(x)$, $\ldots$, $e_k(x)$, we thus also know the results of applying an arbitrary perturbation from the linear space

$$L=\{c_1\cdot e_1(x)+\ldots+c_k\cdot e_k(x)\}.$$

From this viewpoint, it does not matter what exactly perturbations $e_i(x)$ we select as long as they are within the same space L.

Thus, the question of optimally selecting a given number k of perturbations can be formulated as the question of optimally selecting a k-dimensional linear subspace L in the space of all functions.

Shift-invariance: a natural requirement. To select the space L, let us use the fact that in most physical problems, there is no preferred spatial location. Thus, in principle, we can choose different locations as origins ($x=0$) of the coordinate system.

It is reasonable to require that the optimal family of perturbations do not change if we simply change the origin $x=0$. For example, if we select a point with the original coordinates x_0 as the origin of a new coordinate system, then the new coordinates will have the form $x_{\rm new}=x-x_0$. In the original coordinates, the optimal family of perturbations has the form

$$\{c_1\cdot e_1(x)+\ldots+c_k\cdot e_k(x)\}.$$

In the new coordinates $x_{\rm new}$, we should expect the exact same family of perturbations

$$\{c_1\cdot e_1(x_{\rm new})+\ldots+c_k\cdot e_k(x_{\rm new})\}.$$

In terms of the original coordinates, this new family has the form

$$\{c_1 \cdot e_1(x - x_0) + \ldots + c_k \cdot e_k(x - x_0)\}.$$

This "shifted" family must coincide with the original one. In particular, every basis function $e_i(x - x_0)$ from the shifted basis must belong to the original family, i.e., must have the form

$$e_i(x - x_0) = \sum_{j=1}^{k} c_{ij}(x_0) \cdot e_j(x)$$

for some coefficients c_{ij} which are, in general, depending on the shift x_0.

Smoothness: an additional requirement. In many physical problems, it is reasonable to consider smooth perturbations, i.e., perturbations for which the functions $e_i(x)$ are differentiable. In this case, by considering different values x, we get a system of linear equations for determining $c_{ij}(x_0)$ in terms of the smooth functions $e_i(x - x_0)$ and $e_j(x)$. The solution of a system of linear equations is – due to Cramer's rule – a smooth function of the coefficients and of the right-hand sides. Thus, the solutions $c_{ij}(x_0)$ are also smooth.

From the requirements to the description of the desired family L. Let us fix one of the spatial coordinates, e.g., the coordinate x_1. For shifts w.r.t. this coordinate, we have

$$e_i(x_1 - x_0, x_2, \ldots) = \sum_{j=1}^{k} c_{ij}(x_0) \cdot e_j(x_1, x_2, \ldots).$$

Since the functions $e_i(x_1 - x_0, \ldots)$ and $c_{ij}(x_0)$ are smooth, we can differentiate both sides of the above equation with respect to x_0 and take $x_0 = 0$. For each components of x_0, we get a system of linear differential equations

$$e'_i = -\sum c'_{ij}(0) \cdot e_j$$

with constant coefficients. A general solution to such a system is well known: it is a linear combination of expressions $x_1^{k_{1j}} \cdot \exp(a_{1j} \cdot x_1)$ with complex values a_{1j} (eigenvalues of the matrix $-c'_{ij}(0)$) and integers $k_{1j} \geq 0$ (multiplicities of these eigenvalues).

Some of these solutions tend to infinity exponentially fast. Such solutions are not useful as perturbations, since perturbations must be uniformly small. So, it is reasonable to restrict ourselves to *bounded* perturbations.

This boundedness eliminates the terms with $\mathrm{Re}(a_{1j}) \neq 0$. Thus, the only remaining terms correspond to imaginary values a_{1j} – i.e., to sinusoids. For these terms, boundedness also eliminates terms with $k_{1j} > 0$, so we only get pure sinusoids:

$$e_i(x_1, x_2, \ldots) = \sum_j C_j(x_2, \ldots) \cdot \sin(\omega_{1j} \cdot x_1).$$

The functions $C_j(x_2,\ldots)$ can be computed as linear combinations of the values $e_i(x_1,x_2,\ldots)$ corresponding to different values x_1. On the other hand, the dependence of e_i on x_2 is also a linear combination of sinusoids. Thus, the functions $C_j(x_2,\ldots)$ are linear combinations of sinusoids in x_2. Substituting these linear combinations instead of $C_j(x_2,\ldots)$ into the above formula, and taking into account that $\sin(a)\cdot\sin(b)$ is a linear combination of $\cos(a+b)$ and $\cos(a-b)$, we conclude that the dependence of e_i on x_1 and x_2 takes the form

$$e_i(x_1,x_2,x_3,\ldots)=\sum_k C_k(x_3,\ldots)\cdot\sin(\omega_{1k}\cdot x_1+\omega_{2k}\cdot x_2).$$

Similarly, we can add x_3, etc., and conclude that each function $e_i(x)$ is a linear combination of the sinusoids $\sin(\sum\omega_j\cdot x_j+\varphi)$.

Comment. In the above text, we assumed that the desired linear space is shift-invariant. Instead, as we show in the Appendix, we can assume that the desired linear space is optimal with respect to some reasonable optimality criterion – and get the same conclusion, that we should use sinusoids.

Conclusion. We conclude that the optimal perturbations are linear combinations of sinusoids. We thus arrive at the following recommendation: use sinusoidal perturbations.

Acknowledgments.

This work was supported in part by the National Science Foundation grant HRD-0734825, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by Grant MSM 6198898701 from MŠMT of Czech Republic.

The authors are thankful to all the participants of the International Student Conference on Applied Mathematics and Informatics ICAMI'09 (Malenovice, Czech Republic, May 13–15, 2009) for valuable discussions.

References

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

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

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

[4] Kuznetsov, V.: Interval Statistical Methods, Radio i Svyaz Publ. Moscow, 1991 (in Russian).

[5] Nguyen, H. T., Kreinovich, V.: Applications of Continuous Mathematics to Computer Science, Kluwer, Dordrecht, 1997.

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

[7] Perfilieva, I.: Fuzzy transforms: theory and applications, Fuzzy Sets and Systems, 157, 2006, pp. 993–1023.

[8] Perfilieva, I., Novák, V., Dvorák, A.: Fuzzy transform in the analysis of data, International Journal of Approximate Reasoning, 48(1), 2008, pp. 36–46.

[9] Pinheiro da Silva, P., Velasco, A., Ceberio, M., Servin, C., Averill, M., Del Rio, N.: Propagation and provenance of probabilistic and interval uncertainty in cyberinfrastructure-related data processing and data fusion. In: Muhanna, R. L., Mullen, R. L. (eds.), Proceedings of the International Workshop on Reliable Engineering Computing REC'08, Savannah, Georgia, February 20–22, 2008, pp. 199–234.

[10] Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice, Springer Verlag, New York, 2005.

[11] Walley, P.: Statistical Reasoning with Imprecise Probabilities, Wiley, 1991.

A Perturbation Selection as an Optimization Problem

Which linear space is the best? In this Appendix, we consider the problem of selecting the perturbations as the precise optimization problem – and then solve this problem.

Among all possible linear spaces of functions L of a given dimension k, we want to choose the best one. In formalizing what "the best" means we follow the general idea described in [5].

The criteria to choose may be computational simplicity, efficiency of detecting small features, or something else. In mathematical optimization problems, numeric criteria are most frequently used, when to every space we assign some value expressing its performance, and choose a space for which this value is maximal. However, it is not necessary to restrict ourselves to such numeric criteria only. For example,

- if we have several different spaces that have the same (numerically described) ability A to detect small features,
- we can choose, between these spaces, the one that has the minimal computational complexity C.

In this case, the actual criterion that we use to compare two spaces is not numeric, but more complicated: *a space L_1 is better than the family L_2 if and only if either $A(L_1) > A(L_2)$ or $A(L_1) = A(L_2)$ and $C(L_1) < C(L_2)$.*

A criterion can be even more complicated. What a criterion *must* do is to allow us for every pair of spaces L_1 and L_2 to tell whether

- the first space is better with respect to this criterion (we'll denote it by $L_1 > L_2$),
- or the second space is better ($L_1 < L_2$),
- or these spaces have the same quality in the sense of this criterion (we'll denote it by $L_1 \sim L_2$).

Of course, it is necessary to demand that these choices be consistent, e.g., if $L_1 > L_2$ and $L_2 > L_3$ then $L_1 > L_3$.

The criterion must select a unique optimal space. Another natural demand is that this criterion must choose a *unique* optimal space (i.e., a space which is better with respect to this criterion than any other space). The reason for this demand is very simple.

If a criterion does not choose any space at all, then it is of no use.

If several different spaces are "the best" according to this criterion, then we still have a problem to choose among those "best". Therefore, we need some additional criterion for that choice. For example, if several spaces turn out to have the same ability to detect small features, we can choose among them a space with the minimal computational complexity. So what we actually do in this case is abandon that criterion for which there were several "best" spaces, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if

- either it was better according to the old criterion
- or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best space, it means that this criterion is not final; we have to modify it until we come to a final criterion that will have that property.

The criterion must be shift-invariant. The next natural condition that the criterion must satisfy is connected with the shift. Shifting simply means that we change the origin of the coordinate system. Thus, shift should not change the relative quality of the two spaces: if a space L_1 is better than the space L_2, then the "shifted" space $T_{x_0}(L_1) = \{e(x - x_0) : e \in L_1\}$ must be better than a similarly shifted space $T_{x_0}(L_2)$.

Now, we are ready for the formal definitions.

Definition 1. *A pair of relations* $(>, \sim)$ *is called* consistent *if it satisfies the following conditions:*

(1) if $a > b$ and $b > c$ then $a > c$;

(2) $a \sim a$;

(3) if $a \sim b$ then $b \sim a$;

(4) if $a \sim b$ and $b \sim c$ then $a \sim c$;

(5) if $a > b$ and $b \sim c$ then $a > c$;

(6) if $a \sim b$ and $b > c$ then $a > c$;

(7) if $a > b$ then $b > a$ or $a \sim b$ are impossible.

Definition 2. *Assume a set A is given. Its elements will be called* alternatives*.*

- *By an* optimality criterion *we mean a consistent pair $(>, \sim)$ of relations on the set A of all alternatives.*
- *If $a > b$, we say that a is* better *than b.*
- *If $a \sim b$, we say that the alternatives a and b are* equivalent *with respect to this criterion.*
- *We say that an alternative a is* optimal *(or* best*) with respect to a criterion $(>, \sim)$ if for every other alternative b either $a > b$ or $a \sim b$.*

Definition 3. *We say that a criterion is* final *if there exists an optimal alternative, and this optimal alternative is unique.*

In this paper, we fix an integer k, and we consider optimality criteria on the set $\mathcal{L}$ of all k-dimensional linear spaces of bounded smooth (= differentiable) functions.

Definition 4. *Let x_0 be a vector.*

- *By the x_0-*shift *of a function $e(x)$ we mean a function $e_{\text{new}} = T_{x_0}(e)$ for which $e_{\text{new}}(x) = e(x - x_0)$.*
- *By the x_0-*shift *of a space L, we mean a space $T_{x_0}(L) = \{e \mid e \in L\}$.*
- *We say that an optimality criterion on the set $\mathcal{L}$ is* shift-invariant *if for every two spaces L_1 and L_2 and for every vector x_0, the following two conditions are hold:*
  - *i) if L_1 is better than L_2 in the sense of this criterion (i.e., $L_1 > L_2$), then $T_{x_0}(L_1) > T_{x_0}(L_2)$;*
  - *ii) if L_1 is equivalent to L_2 in the sense of this criterion (i.e., $L_1 \sim L_2$), then $T_{x_0}(L_1) \sim T_{x_0}(L_2)$.*

Discussion. As we have already remarked, the demands that the optimality criterion is final and shift-invariant are quite reasonable. The only problem with them is that at first glance they may seem rather weak. However, they are not, as the following theorem shows:

Theorem. *If a space L is optimal in the sense of some optimality criterion that is final and shift-invariant, then all elements of L are linear combinations of sinusoids.*

Proof: first part. Let us first prove that the optimal space $L_{\rm opt}$ exists and is *shift-invariant* in the sense that $L_{\rm opt} = T_{x_0}(L_{\rm opt})$ for all vectors x_0. Indeed, we assumed that the optimality criterion is final, therefore there exists a unique optimal space $L_{\rm opt}$. Let's now prove that this optimal family is shift-invariant.

The fact that $L_{\rm opt}$ is optimal means that for every other space L, either $L_{\rm opt} > L$ or $L_{\rm opt} \sim L$. If $L_{\rm opt} \sim L$ for some $L \neq L_{\rm opt}$, then from the definition of the optimality criterion we can easily deduce that L is also optimal, which contradicts the fact that there is only one optimal family. So for every L,

- either $L_{\rm opt} > L$
- or $L_{\rm opt} = L$.

Let us take an arbitrary x_0 and let $L = T_{x_0}(L_{\rm opt})$. If $L_{\rm opt} > L = T_{x_0}(L_{\rm opt})$, then from the shift-invariance of the optimality criterion (condition ii) we conclude that $T_{-x_0}(L_{\rm opt}) > L_{\rm opt}$, and this conclusion contradicts the choice of $L_{\rm opt}$ as the optimal space. So $L_{\rm opt} > L = T_{x_0}(L_{\rm opt})$ is impossible, and therefore $L_{\rm opt} = L = T_{x_0}(L_{\rm opt})$. Thus, the optimal space is really shift-invariant.

Proof: second part. In the main text, we have already proven that in every shift-invariant linear space of bounded smooth functions, all elements are linear combinations of sinusoids.

So, the theorem is proven.