*axioms*

MDPI

*Article*

# Why Use a Fuzzy Partition in F-Transform?

**Vladik Kreinovich [1,*], Olga Kosheleva [2] and Songsak Sriboonchitta [3]**

1 Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA
2 Department of Teacher Education, University of Texas at El Paso, El Paso, TX 79968, USA
3 Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
* Correspondence: vladik@utep.edu

Received: 21 April 2019; Accepted: 1 August 2019; Published: 2 August 2019

check for updates

**Abstract:** In many application problems, F-transform algorithms are very efficient. In F-transform techniques, we replace the original signal or image with a finite number of weighted averages. The use of a weighted average can be naturally explained, e.g., by the fact that this is what we get anyway when we measure the signal. However, most successful applications of F-transform have an additional not-so-easy-to-explain feature: the fuzzy partition requirement that the sum of all the related weighting functions is a constant. In this paper, we show that this seemingly difficult-to-explain requirement can also be naturally explained in signal-measurement terms: namely, this requirement can be derived from the natural desire to have all the signal values at different moments of time estimated with the same accuracy. This explanation is the main contribution of this paper.

**Keywords:** F-transform; fuzzy partition; measurement; measurement accuracy

---

## 1. Formulation of the Problem

### 1.1. F-Transform: A Brief Reminder

In many practical applications, it turns out to be beneficial to replace the original continuous signal $x(t)$ defined on some time interval with a finite number of "averaged" values

$$x_i = \int A_i(t) \cdot x(t)\, dt, \quad i = 0, \ldots, n, \tag{1}$$

where $A_i(t) \geq 0$ are appropriate functions, see, e.g., [1–6].

In many applications, a very specific form of these functions is used: namely, $A_i(t) = A(t - t_i)$ for some function $A(t)$ and for $t_i = t_0 + i \cdot h$, where $t_0$ and $h > 0$ are numbers for which $A(t)$ is equal to 0 outside the interval $[-h, h]$. However, more general families of functions $A_i(t)$ are also sometimes efficiently used.

The transition from the original function $x(t)$ to the tuple of values $\langle x_0, x_1, \ldots, x_n \rangle$ is known as the *F-transform*; [1–6].

A similar 2-D transformation is very useful in many image processing problems.

### 1.2. The General Idea behind F-Transform Is Very Reasonable

From the general measurement viewpoint, F-transform makes perfect sense because it corresponds to the results of measuring the signal. Indeed, in practice, a measuring instrument cannot measure the exact value $x(t)$ of the signal at a given moment $t$. No matter how fast the processes within the measuring instrument, it always has some inertia. As a result, the value $m_i$ measured at each measurement depends not only on the value $x(t)$ of the signal at the given moment of time, it also depends on the values at nearby moments of time; see, e.g., [7].

The signal is usually weak, so the values $x(t)$ are small. Thus, we can expand the dependence of $m_i$ on $x(t)$ in Taylor series and safely ignore terms which are quadratic or of higher order with respect to $x(t)$. Then, we get a model in which the value $m_i$ is a linear function of different values $x(t)$; this is the usual technique in applications, see, e.g., [8]. The general form of a linear dependence is

$$m_i = m_i^{(0)} + \int A_i(t) \cdot x(t)\, dt \tag{2}$$

for some coefficients $A_i(t)$.

A measuring instrument is usually calibrated in such a way that in the absence of the signal, when $x(t) = 0$, the measurement result is 0. After such a calibration, we get $m_i^{(0)} = 0$ and thus, the expression (2) gets a simplified form

$$m_i = \int A_i(t) \cdot x(t)\, dt. \tag{3}$$

This is exactly the form used in F-transform. Thus, F-transform is indeed a very natural procedure: it replaces the original signal $x(t)$ with the simulated results of measuring this signal, and the results of measuring the signal is exactly what we have in real life.

### 1.3. However, Why a Fuzzy Partition?

So far, everything has been good and natural, but there is one aspect of successful applications of F-transform that cannot be explained so easily: namely, that in most such applications, the corresponding functions $A_i(t)$ form a *fuzzy partition*, in the sense that

$$\sum_{i=1}^{n} A_i(t) = 1 \tag{4}$$

for all moments $t$ from the corresponding time interval.

#### 1.3.1. Mathematical Comment

Sometimes, the corresponding requirement takes a slightly different form $\sum\limits_{i=1}^{n} A_i(t) = c$ for some constant $c$. This case can be naturally reduced to case (4) if we consider re-scaled functions $A'_i(t) = c^{-1} \cdot A_i(t)$ and the corresponding re-scaled values $x'_i = c^{-1} \cdot x_i$. In view of this equivalent re-scaling, the question is why it is natural to require that $\sum\limits_{i=1}^{n} A_i(t) = \text{const}$.

#### 1.3.2. Application-Related Comment

It is worth mentioning that fuzzy partitions are successfully used in other applications of fuzzy techniques. For example, fuzzy sets that form a fuzzy partition are used:

- in fuzzy control; see, e.g., an application to control of telerobots in space medicine [9];
- in information accessing systems such as information retrieval systems, filtering systems, recommender systems, and web quality evaluation tools; see, e.g., [10] and references therein, etc.

In fuzzy clustering, the important frequently used requirement is also that the fuzzy sets corresponding to different clusters form a fuzzy partition. The resulting clustering techniques have been very successful in many applications; see, e.g., a recent application to the analysis of earthquake data [11].

On the other hand, in some other applications, it turns to be more efficient to use fuzzy sets which *do not* form a fuzzy partition; an example related to face and pose detection is given in [12].

### *1.4. It Is Desirable to Explain the Efficiency of a Fuzzy Partition Requirement*

We strongly believe that every time there is an unexplained empirical fact about data processing algorithms, it is desirable to come up with a theoretical explanation. Such an explanation makes the resulting algorithms more reliable, thus decreasing the possibility that these algorithms will fail and, correspondingly, increasing the chances that these efficient algorithms will be used by practitioners, even in potentially high-risk situations. Sometimes, the corresponding analysis finds conditions under which these methods work efficiently, and even helps develop even more efficient techniques.

In applications of fuzzy logic, there are many such interesting empirical facts, e.g., higher efficiency of certain membership functions, of certain "and" and "or"-operations, of certain defuzzification procedures, etc. Finding an explanation for such facts has been one of our main research directions; see, e.g., [13–27]. Similar research results help understand the somewhat unexpected efficiency of other intelligent techniques such as deep neural networks; see, e.g., [28–33]. This paper can be viewed as a natural continuation of this research direction.

### *1.5. What We Do in this Paper*

In this paper, we show that the fuzzy partition requirement (4) can be naturally explained in the measurement interpretation of F-transform.

To be more precise, we show that what naturally appears is a 1-parametric family of similar requirements of which the fuzzy partition requirement is a particular case, and then we explain that in the fuzzy cases, it is indeed reasonable to use the fuzzy partition requirement.

The resulting explanation of the fuzzy partition requirement is the main contribution of this paper.

### *1.6. The Structure of this Paper*

The structure of our paper is as follows. The main idea behind our explanation is presented in Section 2, for a very general (not necessary fuzzy) uncertainty. In Section 3, we analyze the case of probabilistic uncertainty. In Section 4, this analysis is generalized to general uncertainty. In Section 5, the analysis performed in the previous sections is used to explain which functions $A_i(t)$ we should choose in a general uncertainty situation and, in particular, in the case of fuzzy uncertainty. All these sections contain original research results. The final Section 6 contains conclusions and possible future research directions.

#### Comment

In the above text, we assumed that the actual signal $x(t)$ is defined for all possible moments of time. In some practical situations, however, it makes sense to only consider discrete-time values $x(t_1)$ $x(t_2), \ldots$ In such discrete-time situations, it makes sense to apply the following *discrete F-transform*:

$$x_i = c \cdot \sum_k A_i(t_k) \cdot x(t_k)$$

for some constant $c$, where the values $A_i(t_k)$ satisfy the same fuzzy partition requirement: that for each $k$, we have $\sum\limits_i A_i(t_k) = 1$.

In the following text, we use the continuous case to explain the fuzzy partition requirement; however, as one can easily check, the same explanation holds for the discrete F-transform as well.

## 2. Main Idea

### *2.1. What if We Can Make Exact Measurements of Instantaneous Values?*

In the idealized case, when inertia of measuring instruments is so small that it can be safely ignored, we can measure the exact values $x(t_1)$, $x(t_2)$, $\ldots$, of the signal $x(t)$ at different moments of time.

In this case, we get perfect information about the values of the signal at these moments of time $t_1$, $t_2, \ldots$, but practically no information about the values of the signal $x(t)$ at any other moment of time. In other words:

- we reconstruct the values $x(t_1)$, $x(t_2)$, $\ldots$, with perfect accuracy (0 measurement error), while
- the values $x(t)$ corresponding to all other moments of time $t$ are reconstructed with no accuracy at all (the only bound on measurement error is infinity).

Even if we take into account that measurements are never 100% accurate, and we only measure the values $x(t_i)$ with some accuracy, we will still get the difference between our knowledge of values $x(t)$ corresponding to different moments of time:

- we know the values $x(t_i)$ with finite accuracy, but
- for all other moments of time $t$, we know nothing (i.e., the only bound of measurement error is infinity).

This difference does not fit well with the fact that we want to get a good representation of the *whole* signal $x(t)$, i.e., a good representation of its values at all moments of time. Thus, we arrive at the following idea.

### *2.2. Main Idea*

To adequately represent the original signal $x(t)$, it is desirable to select the measurement procedures in such a way that based on these measurements, we reconstruct each value $x(t)$ with the same accuracy.

Comment

At this moment, we have presented this idea informally. In the following sections, we will show how to formalize this idea, and we also show that this idea leads to the fuzzy partition requirement.

To be more precise, this idea leads to a general formula that includes the fuzzy partition requirement as a particular case. We also explain why namely the fuzzy partition requirement should be selected in the fuzzy case.

## 3. Case of Probabilistic Uncertainty

### *3.1. Description of the Case*

Let us start with the most well-studied uncertainty: the probabilistic uncertainty. In this case, we have probabilistic information about the measurement error $\Delta m_i \stackrel{\text{def}}{=} \widetilde{m}_i - m_i$ of each measurement, where $\widetilde{m}_i$ denotes the result of measuring the quantity $m_i$.

We will consider the usual way measurement uncertainties are treated in this approach (see, e.g., [7]): namely, we will assume:

- that each measurement error $\Delta m_i$ is normally distributed with 0 mean and known standard deviation $\sigma$, and
- that measurement errors $\Delta m_i$ and $\Delta m_j$ corresponding to different measurements $i \neq j$ are independent.

*3.2. How Accurately Can We Estimate $X(T)$ Based on Each Measurement*

Based on each measurement, we know each value $m_i = \int A_i(t) \cdot x(t)\,dt$ with accuracy $\sigma$. The integral is, in effect, a large sum, so we have

$$m_i = \sum_t A_i(t) \cdot x(t) \cdot \Delta t.$$

Thus, for each moment $t$, we have

$$A_i(t) \cdot x(t) \cdot \Delta t = m_i - \sum_{s \neq t} A_i(s) \cdot x(s) \cdot \Delta s, \tag{5}$$

and therefore,

$$x(t) = \frac{1}{A_i(t) \cdot \Delta t} \cdot m_i - \frac{1}{A_i(t) \cdot \Delta t} \cdot m_i \cdot \sum_{s \neq t} A_i(s) \cdot x(s) \cdot \Delta s. \tag{6}$$

The measurement result $\widetilde{m}_i$ is an estimate for the quantity $m_i$, with mean 0 and standard deviation $\sigma$. Thus, if we know all the values $x(s)$ corresponding to $s \neq t$, then, based on the result $\widetilde{m}_i$ of the $i$-th measurement, we can estimate the remaining value $x(t)$ as

$$x(t) \approx \widetilde{x}_i(t) \stackrel{\text{def}}{=} \frac{1}{A_i(t) \cdot \Delta t} \cdot \widetilde{m}_i - \frac{1}{A_i(t) \cdot \Delta t} \cdot m_i \cdot \sum_{s \neq t} A_i(s) \cdot x(s) \cdot \Delta s. \tag{7}$$

By comparing the Formulas (6) and (7), we can conclude that the approximation error $\Delta x_i(t) \stackrel{\text{def}}{=} \widetilde{x}_i(t) - x(t)$ of this estimate is equal to

$$\Delta x_i(t) = \frac{1}{A_i(t) \cdot \Delta t} \cdot \Delta m_i. \tag{8}$$

Since the measurement error $\Delta m_i$ is normally distributed, with 0 mean and standard deviation $\sigma$, the approximation error $\Delta x_i(t)$ is also normally distributed, with 0 mean and standard deviation

$$\sigma_i(t) = \frac{\sigma}{A_i(t) \cdot \Delta t}. \tag{9}$$

*3.3. How Accurately Can We Estimate $X(T)$ Based on All The Measurements*

For each moment $t$, based on each measurement $i$, we get an estimate $\widetilde{x}_i(t) \approx x(t)$ with the accuracy $\sigma_i$ described by Formula (9):

$$\begin{aligned} x(t) &\approx \widetilde{x}_0(t), \\ x(t) &\approx \widetilde{x}_1(t), \\ &\ldots \\ x(t) &\approx \widetilde{x}_n(t). \end{aligned} \tag{10}$$

For each estimate, since the distribution of the measurement error is normal, the corresponding probability density function has the form

$$\rho_i(\widetilde{x}_i(t)) = \frac{1}{\sqrt{\pi} \cdot \sigma_i(t)} \cdot \exp\left(-\frac{(\widetilde{x}_i(t) - x(t))^2}{2(\sigma_i(t))^2}\right). \tag{11}$$

Since the measurement errors $\Delta m_i$ of different measurements are independent, the resulting estimation errors $\Delta x_i(t) = \widetilde{x}_i(t) - x(t)$ are also independent. Thus, the joint probability density corresponding to all the measurements is equal to the product of all the values (11) corresponding to individual measurements:

$$\rho(\widetilde{x}_0(t),\ldots,\widetilde{x}_n(t)) = \frac{1}{(\sqrt{\pi})^{n+1}\cdot\prod\limits_{i=0}^{n}\sigma_i(t)}\cdot\exp\left(-\sum_{i=0}^{n}\frac{(\widetilde{x}_i(t)-x(t))^2}{2(\sigma_i(t))^2}\right). \tag{12}$$

As a combined estimate $\widetilde{x}(t)$ for $x(t)$, it is reasonable to select the value for which the corresponding probability (12) is the largest possible. This is known as the *Maximum Likelihood Method*; see, e.g., [34].

To find such a maximum, it is convenient to take the negative logarithm of expression (12) and use the fact that $-\ln(z)$ is a decreasing function, so the original expression is the largest if and only if its negative logarithm is the smallest. Thus, we arrive at the need to minimize the sum

$$\sum_{i=0}^{n}\frac{(\widetilde{x}_i(t)-x(t))^2}{2\sigma_i^2}; \tag{13}$$

this minimization is known as the *Least Squares* approach.

Differentiating expression (13) with respect to the unknown $x(t)$ and equating the derivative to 0, we conclude that

$$\sum_{i=0}^{n}\widetilde{x}_i(t)\cdot(\sigma_i(t))^{-2} = \widetilde{x}(t)\cdot\sum_{i=0}^{n}(\sigma_i(t))^{-2}, \tag{14}$$

and thus that

$$\widetilde{x}(t) = \frac{\sum\limits_{i=0}^{n}\widetilde{x}_i(t)\cdot(\sigma_i(t))^{-2}}{\sum\limits_{i=0}^{n}(\sigma_i(t))^{-2}}. \tag{15}$$

The accuracy $\widetilde{\sigma}(t)$ of this estimate can be determined if we describe the expression (12) in the form

$$\frac{1}{\sqrt{\pi}\cdot\widetilde{\sigma}(t)}\cdot\exp\left(-\frac{(x(t)-\widetilde{x}(t))^2}{2(\widetilde{\sigma}(t))^2}\right). \tag{16}$$

By comparing the coefficients at $(x(t))^2$ under the exponent in Formulas (12) and (16), we conclude that

$$\frac{1}{2(\widetilde{\sigma}(t))^2} = \sum_{i=0}^{n}\frac{1}{2(\sigma_i(t))^2}, \tag{17}$$

i.e., equivalently that

$$(\widetilde{\sigma}(t))^{-2} = \sum_{i=0}^{n}(\sigma_i(t))^{-2}. \tag{18}$$

In particular, if all the estimation errors were equal, i.e., if we had $\sigma_i(t) = \sigma(t)$ for all $i$, then, from (18), we would conclude that

$$\widetilde{\sigma}(t) = \frac{\sigma(t)}{\sqrt{N}}, \tag{19}$$

where $N \stackrel{\text{def}}{=} n+1$ is the overall number of combined measurements.

Substituting expression (9) for $\sigma_i(t)$ into Formula (18), we conclude that

$$(\widetilde{\sigma}(t))^{-2} = \frac{(\Delta t)^2}{\sigma^2}\cdot\sum_{i=0}^{n}(A_i(t))^2. \tag{20}$$

Thus, the requirement that we get the same accuracy for all moments of time $t$, i.e., that $\widetilde{\sigma}(t) =$ const means that we need to have

$$\sum_{i=0}^{n} (A_i(t))^2 = \text{const}. \tag{21}$$

### 3.4. Discussion

Formula (21) is somewhat similar to the fuzzy partition requirement but it is different:

- in the fuzzy partition requirement, we demand that the sum of the functions $A_i(t)$ be constant, but
- here, we have the sum of the squares.

Formula (21) is based on the probabilistic uncertainty, for which the measurement error decreases with repeated measurements as $1/\sqrt{N}$. However, e.g., for interval uncertainty (see, e.g., [7,35–37]), when we only know the upper bound on the measurement errors, the measurement error resulting from $N$ repeated measurements decreases as $1/N$; see, e.g., [38].

So maybe by considering different types of uncertainty, we can get the fuzzy partition formula? To answer this question, let us consider a general case of how uncertainties can be combined in different approaches.

## 4. How Uncertainties Can Be Combined in Different Approaches

### 4.1. Towards a General Formulation of the Problem

In the general case, be it probabilistic or interval or any other approach, we can always describe the corresponding uncertainty in the same unit as the measured quantity.

In the interval approach, a natural measure of uncertainty is the largest possible value $\Delta$ of the absolute value $|\Delta x|$ of the approximation error $\Delta x = \widetilde{x} - x$, where $x$ is the actual value of the corresponding quantity and $\widetilde{x}$ is the measurement result. This value $\Delta$ is clearly measured in the same units as the quantity $x$ itself.

In the probabilistic approach, we can use the variance of $\Delta x$—which is described in different units than $x$—but we can also take the square root of this variance and consider standard deviation $\sigma$, which is already described in the same units.

In the general case, let us denote the corresponding measure of accuracy by $\Delta$. The situation when we have no information about the desired quantity corresponds to $\Delta = \infty$. The idealized situation when we know the exact value of this quantity corresponds to $\Delta = 0$.

If $\Delta$ and $\Delta'$ are corresponding measures of accuracy for two different measurements, then what is the accuracy of the resulting combined estimate? Let us denote this combined accuracy by $\Delta * \Delta'$.

In these terms, to describe the combination, we need to describe the corresponding function $a * b$ of two variables. What are the natural properties of this function?

### 4.2. Commutativity

The result of combining two estimates should not depend on which of the two estimates is listed first, so we should have $a * b = b * a$. In other words, the corresponding combination operation must be commutative.

### 4.3. Associativity

If we have three estimates, then:

- we can first combine the first and the second ones, and then combine the result with the third one,
- or we can first combine the second and the third ones, and then combine the result with the first one.

The result should not depend on the order, so we should have $(a * b) * c = a * (b * c)$. In other words, the corresponding operation should be associative.

### 4.4. Monotonicity

Any additional information can only improve the accuracy. Thus, the accuracy of the combined estimate cannot be worse than the accuracy of each of the estimates used in this combination. Therefore, we get $a * b \leq a$.

Similarly, if we increase the accuracy of each measurement, the accuracy of the resulting measurement will increase too: if $a \leq a'$ and $b \leq b'$, then we should have $a * b \leq a' * b'$.

### 4.5. Non-Degenerate Case

If we start with measurements of finite accuracy, we should never get the exact value, i.e., if $a > 0$ and $b > 0$, we should get $a * b > 0$.

### 4.6. Scale-Invariance

In real life, we deal with the actual quantities, but in computations, we need to describe these quantities by their numerical values. To get a numerical value, we need to select a measuring unit: e.g., to describe distance in numerical terms, we need to select a unit of distance.

This selection is usually arbitrary. For example, for distance, we could consider meters, we could consider centimeters, and we could consider inches or feet. It is reasonable to require that the combination operation remains the same if we keep the same quantities but change the measuring unit. Let us describe this requirement in precise terms.

If we replace the original measuring unit with a new one which is $\lambda$ times smaller, then all the numerical values are multiplied by $\lambda$. For example, if we replace meters by centimeters, then all the numerical values are multiplied by 100. The corresponding transformation $x \to \lambda \cdot x$ is known as *scaling*.

Suppose that in the original units, we had accuracies $a$ and $b$ and the combined accuracy was $a * b$. Then, in the new units—since accuracies are described in the same units as the quantity itself—the original accuracies become $\lambda \cdot a$ and $\lambda \cdot b$, and the combined accuracy is thus $(\lambda \cdot a) * (\lambda \cdot b)$. This is the combined accuracy in the new units. It should be the same as when we transform the old-units accuracy $c = a * b$ into the new units, getting $\lambda \cdot (a * b)$: $(\lambda \cdot a) * (\lambda \cdot b) = \lambda \cdot (a * b)$. This invariance under scaling is known as *scale-invariance.*

### 4.7. Discussion

Now, we are ready to formulate the main result. To formulate it, we list all the above reasonable properties of a combination operation in the form of the following definition:

**Definition 1.** By a *combination operation*, we mean a function $a * b$ that transforms two non-negative numbers $a$ and $b$ into a new non-negative number and for which the following properties hold:

- for all $a$ and $b$, we have $a * b = b * a$ (*commutativity*);
- for all $a$, $b$, and $c$, we have $(a * b) * c = a * (b * c)$ (*associativity*);
- for all $a$ and $b$, we have $a * b \leq a$ (*first monotonicity requirement*);
- for all $a$, $b$, $a'$, and $b'$, if $a \leq a'$ and $b \leq b'$, then $a * b \leq a' * b'$ (*second monotonicity requirement*);
- if $a > 0$ and $b > 0$, then $a * b > 0$ (*non-degeneracy*); and
- for all $a$, $b$, and $\lambda > 0$, we have $(\lambda \cdot a) * (\lambda \cdot b) = \lambda \cdot (a * b)$ (*scale-invariance*).

Comment

This definition is similar to similar definitions presented in [39] for quantum systems and in [30] for the neural networks. However, because of the different application domains, the above definition is somewhat different: e.g., in our case, we have non-degeneracy requirement which is natural for combining uncertainty but not in the above two domains.

**Proposition 1.** Every combination operation has either the form $a * b = \min(a, b)$ or the form $a * b = (a^{-\beta} + b^{-\beta})^{-1/\beta}$ for some $\beta > 0$.

*Proof* of this result is, in effect, described in [39] (see also [30]).

Comment

The proof shows that if we do not impose the non-degeneracy condition, the only other alternative is $a * b = 0$. Thus, the non-degeneracy condition can be weakened: instead of requiring that $a * b > 0$ for *all* pairs of positive numbers $a$ and $b$, it is sufficient to require that $a * b > 0$ for at least *one* such pair.

### *4.8. Discussion*

The form $\min(a, b)$ is the limit case of the second form when $\beta \to \infty$.

In the generic case $\beta < \infty$, $a * b = c$ is equivalent to

$$a^{-\beta} + b^{-\beta} = c^{-\beta}. \tag{22}$$

Thus, the probabilistic case corresponds to $\beta = 2$.

In the situation when we have $N$ measurement results with the same accuracy $\Delta_1 = \ldots = \Delta_N = \Delta$, the combined accuracy $\widetilde{\Delta}$ can be determined from the formula

$$\left(\widetilde{\Delta}\right)^{-\beta} = N \cdot \Delta^{-1/\beta}. \tag{23}$$

Thus, we have

$$\widetilde{\Delta} = \frac{\Delta}{N^{1/\beta}}. \tag{24}$$

In the probabilistic case, we indeed have this formula with $\beta = 2$. The above-mentioned interval-case formula $\widetilde{\Delta} \sim \frac{1}{N}$ (derived in [38]) corresponds to the case $\beta = 1$; thus, $\beta = 1$ is the value of the parameter $\beta$ corresponding to interval uncertainty.

## 5. Which Functions $A_i(T)$ Should We Choose: General Uncertainty Situation and Case of Fuzzy Uncertainty

### *5.1. Analysis of the Problem*

If we measure $m_i$ with accuracy $\Delta$, then, due to Formula (8) (and similarly to the case of probabilistic uncertainty), the estimate $\widetilde{x}_i(t)$ is known with accuracy

$$\Delta_i(t) = \frac{\Delta}{A_i(t) \cdot \Delta t}. \tag{25}$$

For the case of min combination formula, the combined accuracy is equal to

$$\Delta(t) = \min_i \Delta_i(t) = \frac{\Delta}{\Delta t} \cdot \frac{1}{\max\limits_i A_i(t)}. \tag{26}$$

Thus, the requirement that we estimate all the values $x(t)$ with the same accuracy means that

$$\max_i A_i(t) = \text{const}. \tag{27}$$

For the generic case $\beta < \infty$, from Formula (22), we conclude that

$$(\Delta(t))^{-\beta} = \frac{(\Delta t)^{\beta}}{\Delta^{\beta}} \cdot \sum_{i=1}^{n} (A_i(t))^{\beta}. \tag{28}$$

Thus, the requirement that we get the same accuracy for all moments of time $t$ means that we need to have

$$\sum_{i=1}^{n}(A_i(t))^{\beta} = \text{const.} \tag{29}$$

### 5.2. General Conclusion

The requirement that we get the same accuracy for reconstructing the value of the signal at each moment of time of $t$ leads either to condition (27) or to condition (29). In particular, for $\beta = 1$, we get the fuzzy partition property.

### 5.3. Which Value $\beta$ Should We Use in the Case of Fuzzy Uncertainty

In the fuzzy case (see, e.g., [40–48]), the usual way of propagating uncertainty—Zadeh extension principle—is equivalent to applying interval computations for each $\alpha$-cut. Thus, for analyzing fuzzy data, it makes sense to use the value of $\beta$ corresponding to interval uncertainty, which as we have mentioned at the end of the previous section, is $\beta = 1$. For $\beta = 1$, Formula (29) becomes the fuzzy partition property. Thus, *when analyzing fuzzy data, the use of fuzzy partition property is indeed justified.*

## 6. Conclusions and Future Work

### 6.1. Conclusions

In many applications of fuzzy techniques, including applications of F-transforms, we use fuzzy sets $A_1(t), \ldots, A_n(t)$ that form a *fuzzy partition*, in the sense that for each $t$, the corresponding degrees $A_i(t)$ add up to 1 (or to a constant): $\sum\limits_i A(t) = 1$. Empirically, in many applications, the fuzzy partition requirement indeed helps, but why does it help? This, until now, remained a mystery.

In this paper, we provide a theoretical justification for this requirement. Specifically, we show that the fuzzy partition requirement naturally follows from the desire to have the signal values at different moments of time to be estimated with the same accuracy.

### 6.2. Possible Directions of Future Research

While our main objective was to explain the ubiquity of the *fuzzy partition* requirement in fuzzy logic, our analysis started on a more general note, by considering *general uncertainty*, of which fuzzy is a particular case. In addition to the case of fuzzy uncertainty, we also explicitly analyzed another important particular type of uncertainty: probabilistic uncertainty.

It is desirable to extend this analysis to other types of uncertainty, e.g.,:

- to different imprecise probability situations, and
- to situations when different functions $A_i(t)$ correspond to different types of uncertainty.

It is also desirable to analyze the situations (like the situation mentioned in Section 1) when empirically, fuzzy sets that *do not* form a fuzzy partition work better. Maybe in this case, a more general scheme with $\beta \neq 1$ will help?

**Author Contributions:** The authors contributed equally to this work.

**Funding:** This work was supported by the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand, and by the US National Science Foundation via grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

**Acknowledgments:** The authors are thankful to Irina Perfilieva for encouragement and valuable discussions, and to the anonymous referees for very useful suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. Novak, V.; Perfilieva, I.; Holcapek, M.; Kreinovich, V. Filtering out high frequencies in time series using F-transform. *Inf. Sci.* **2014**, *274*, 192–209. [CrossRef]
2. Novak, V.; Perfilieva, I.; Kreinovich, V. F-transform in the analysis of periodic signals. In Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012, Osaka, Japan, 24–27 September 2012.
3. Perfilieva, I. Fuzzy transforms: Theory and applications. *Fuzzy Sets Syst.* **2006**, *157*, 993–1023. [CrossRef]
4. Perfilieva, I. F-transform. In *Springer Handbook of Computational Intelligence*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 113–130.
5. Perfilieva, I.; Danková, M.; Bede, B. Towards a higher degree F-transform. *Fuzzy Sets Syst.* **2011**, *180*, 3–19. [CrossRef]
6. Perfilieva, I.; Kreinovich, V.; Novak, V. F-transform in view of trend extraction. In Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012, Osaka, Japan, 24–27 September 2012.
7. Rabinovich, S.G. *Measurement Errors and Uncertainties: Theory and Practice*; Springer: New York, NY, USA, 2005.
8. Feynman, R.; Leighton, R.; Sands, M. *The Feynman Lectures on Physics*; Addison Wesley: Boston, MA, USA, 2005.
9. Haidegger, T.; Kovácz, L.; Precup, R.-E.; Benýo, B.; Benyó, Z.; Preitl, S. Simulation and control in telerobots in space medicine. *Acta Astronaut.* **2012**, *81*, 390–402. [CrossRef]
10. Herrera-Viedma, E.; López-Herrera, A.G. A review on information accessing systems based on fuzzy linguistic modeling. *Int. J. Comput. Intell. Syst.* **2010**, *3*, 420–437. [CrossRef]
11. Di Martino, F.; Pedrycz, W.; Sessa, S. Spatiotemporal extended fuzzy C-means clustering algorithm for hotspots detection and prediction. *Fuzzy Sets Syst.* **2018**, *340*, 109–126. [CrossRef]
12. Moallem, P.; Mousavi, B.S.; Naghibzadeh, S.S. Fuzzy inference system optimized by genetic algorithm for robust face and pose detection. *Int. J. Artif. Intell.* **2015**, *13*, 73–88.
13. Kosheleva, O.; Kreinovich, V. Why Ragin's fuzzy techniques lead to successful social science applications: An explanation. *J. Innov. Technol. Educ.* **2016**, *3*, 185–192. [CrossRef]
14. Kosheleva, O.; Kreinovich, V. Why product "and?"-operation is often efficient: one more argument. *J. Innov. Technol. Educ.* **2017**, *4*, 25–28. [CrossRef]
15. Kosheleva, O.; Kreinovich, V. Why Bellman-Zadeh approach to fuzzy optimization. *Appl. Math. Sci.* **2018**, *12*, 517–522. [CrossRef]
16. Kosheleva, O.; Kreinovich, V.; Ngamsantivong, T. Why complex-valued fuzzy? Why complex values in general? A computational explanation. In Proceedings of the Joint World Congress of the International Fuzzy Systems Association and Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS' 2013, Edmonton, AB, Canada, 24–28 June 2013; pp. 1233–1236.
17. Kreinovich, V. Why intervals? Why fuzzy numbers? Towards a new justification. In Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence, Honolulu, HI, USA, 1–5 April 2007.
18. Kreinovich, V.; Mouzouris, G.; Nguyen, H.T. Fuzzy rule based modeling as a universal approximation tool. In *Fuzzy Systems: Modeling and Control*; Nguyen, N.T., Sugeno, M., Eds.; Kluwer: Boston, MA, USA, 1998; pp. 135–195.
19. Kreinovich, V.; Perfilieva, I.; Novak, V. Why inverse F-transform? A compression-based explanation. In Proceedings of the 2013 International Conference on Fuzzy Systems FUZZ-IEEE'2013, Hyderabad, India, 7–10 July 2013; pp. 1378–1384.
20. Kreinovich, V.; Stylios, C. Why fuzzy cognitive maps are efficient. *Int. J. Comput. Commun. Control* **2015**, *10*, 825–833. [CrossRef]
21. Kreinovich, V.; Stylios, C. when should we switch from interval-valued fuzzy to full type-2 fuzzy (e.g., Gaussian)? *Crit. Rev.* **2015**, *XI*, 57–66.
22. Nguyen, H.T.; Koshelev, M.; Kosheleva, O.; Kreinovich, V.; Mesiar, R. Computational complexity and feasibility of fuzzy data processing: why fuzzy numbers, which fuzzy numbers, which operations with fuzzy numbers. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98), Paris, France, 6–10 July 1998; pp. 273–280.

23. Nguyen, H.T.; Kreinovich, V. *Applications of Continuous Mathematics to Computer Science*; Kluwer: Dordrecht, The Netherlands, 1997.
24. Nguyen, H.T.; Kreinovich, V. Methodology of fuzzy control: An introduction. In *Fuzzy Systems: Modeling and Control*; Nguyen, H.T., Sugeno, M., Eds.; Kluwer: Boston, MA, USA, 1998; pp. 19–62.
25. Nguyen, H.T.; Kreinovich, V.; Lorkowski, J.; Abu, S. Why Sugeno lambda-measures. In Proceedings of the IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2015, Istanbul, Turkey, 1–5 August 2015.
26. Ouncharoen, R.; Kreinovich, V.; Nguyen, H.T. Why lattice-valued fuzzy values? A mathematical justification. *J. Intell. Fuzzy Syst.* **2015**, *29*, 1421–1425. [CrossRef]
27. Perfilieva, I.; Kreinovich, V. Why fuzzy transform is efficient in large-scale prediction problems: A theoretical explanation. *Adv. Fuzzy Syst.* **2011**, *2011*. [CrossRef]
28. Baral, C.; Fuentes, O.; Kreinovich, V. Why deep neural networks: A possible theoretical explanatio. In *Constraint Programming and Decision Making: Theory and Applications*; Ceberio, M., Kreinovich, V., Eds.; Springer Verlag: Berlin/Heidelberg, Germany, 2018; pp. 1–6.
29. Farhan, A.; Kosheleva, O.; Kreinovich, V. Why max and average poolings are optimal in convolutional neural networks, In Proceedings of the 7th International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making IUKM' 2019, Nara, Japan, 27–29 March 2019.
30. Gholamy, A.; Parra, J.; Kreinovich, V.; Fuentes, O.; Anthony, E. How to best apply deep neural networks in geosciences: Towards optimal 'averaging' in dropout training. In *Smart Unconventional Modelling, Simulation and Optimization for Geosciences and Petroleum Engineering*; Watada, J., Tan, S.C., Vasant, P., Padmanabhan, E., Jain, L.C., Eds.; Springer: Amsterdam, The Netherlands, 2019; pp. 15–26.
31. Kosheleva, O.; Kreinovich, V. Why deep learning methods use KL divergence instead of least squares: A possible pedagogical explanation. *Math. Struct. Simul.* **2018**, *46*, 102–106.
32. Kreinovich, V. From traditional neural networks to deep learning: Towards mathematical foundations of empirical successes. In Proceedings of the World Conference on Soft Computing, Baku, Azerbaijan, 29–31 May 2018.
33. Nava, J.; Kreinovich, V. Why a model produced by training a neural network is often more computationally efficient than a nonlinear regression model: A theoretical explanation. *J. Uncertain Syst.* **2014**, *8*, 193–204.
34. Sheskin, D.J. *Handbook of Parametric and Nonparametric Statistical Procedures*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2011.
35. Jaulin, L.; Kiefer, M.; Didrit, O.; Walter, E. *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*; Springer: London, UK, 2001.
36. Mayer, G. *Interval Analysis and Automatic Result Verification*; De Gruyter: Berlin, Germany, 2017.
37. Moore, R.E.; Kearfott, R.B.; Cloud, M.J. *Introduction to Interval Analysis*; SIAM: Philadelphia, PA, USA, 2009.
38. Walster, G.W.; Kreinovich, V. For unknown–but–bounded errors, interval estimates are often better than averaging. *ACM SIGNUM Newsl.* **1996**, *31*, 6–19. [CrossRef]
39. Autchariyapanitkul, A.; Kosheleva, O.; Kreinovich, V.; Sriboonchitta, S. Quantum econometrics: How to explain its quantitative successes and how the resulting formulas are related to scale invariance, entropy, and fuzziness. In Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making IUKM' 2018, Hanoi, Vietnam, 13–15 March 2018; Huynh, V.-N., Inuiguchi, M., Tran, D.-H., Denoeux, T., Eds.; Springer: Cham, Switzerland, 2018.
40. Belohlavek, R.; Dauben, J.W.; Klir, G.J. *Fuzzy Logic and Mathematics: A Historical Perspective*; Oxford University Press: New York, NY, USA, 2017.
41. Klir, G.; Yuan, B. *Fuzzy Sets and Fuzzy Logic*, Prentice Hall: Upper Saddle River, NJ, USA, 1995.
42. Mendel, J.M. *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*; Springer: Cham, Switzerland, 2017.
43. Nguyen, H.T.; Kreinovich, V. Nested intervals and sets: Concepts, relations to fuzzy sets, and applications. In *Applications of Interval Computations*; Kearfott, R.B., Kreinovich, V., Eds.; Kluwer: Dordrecht, The Netherlands, 1996; pp. 245–290.
44. Nguyen, H.T.; Walker, C.; Walker, E.A. *A First Course in Fuzzy Logic*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.

45. Novák, V.; Perfilieva, I.; Močkoř, J. *Mathematical Principles of Fuzzy Logic*; Kluwer: Boston, MA, USA; Dordrecht, The Netherlands, 1999.
46. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning—I. *Inf. Sci.* **1975**, *8*, 199–249. [CrossRef]
47. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning—II. *Inf. Sci.* **1975**, *8*, 301–357. [CrossRef]
48. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning—III. *Inf. Sci.* **1975**, *9*, 43–80. [CrossRef]