

# Why Triangular Membership Functions Are Successfully Used in F-Transform Applications: A Global Explanation to Supplement the Existing Local Ones

Olga Kosheleva<sup>1</sup>, Vladik Kreinovich<sup>1</sup>, and Thach Ngoc Nguyen<sup>2</sup>

<sup>1</sup>University of Texas at El Paso

El Paso, Texas 79968, USA

olgak@utep.edu, vladik@utep.edu

<sup>2</sup>Banking University of Ho Chi Minh City

56 Hoang Dieu 2, Quan Thu Duc, Thu Duc

Ho Chi Minh City, Vietnam, Thachnn@buh.edu.vn

## Abstract

The main ideas of F-transform came from representing expert rules. It would be therefore reasonable to expect that the more accurately the membership functions describe human reasoning, the more successful will be the corresponding F-transform formulas. We know that an adequate description of our reasoning corresponds to complicated membership functions – however, somewhat surprisingly, many successful applications of F-transform use the simplest possible triangular membership functions. There exist some explanations for this phenomenon, which are based on local behavior of the signal. In this paper, we supplement these local explanations by a global one: namely, we prove that triangular membership functions are the only one that provide the exact reconstruction of the appropriate global characteristic of the signal.

**Keywords:** F-transform; triangular membership function; optimal global characteristic

## 1 Formulation of the Problem

**F-transforms: a brief reminder.** In many application areas, it turned to be very useful to transform the original signal  $x(t)$  into the values proportional to

$$x_i = \int A\left(\frac{t-t_i}{h}\right) \cdot x(t) dt, \quad (1)$$

where  $t_i = t_0 + i \cdot h$  for appropriate  $t_0$  and  $h > 0$ , and  $A(t)$  is a non-negative function:

- which is equal to 0 outside the interval  $[-1, 1]$ ,
- which, starting at  $t = -1$ , increases to 1 until it reaches  $t = 0$ ,
- which then decreases to 0, and
- for which

$$\sum_i A\left(\frac{t-t_i}{h}\right) = 1 \quad (2)$$

for all  $t$ ; this last property is known as the *fuzzy partition property*.

This transformation is known as *F-transform*; see, e.g., [14, 15, 17, 18, 19, 20].

This transform comes from the general fuzzy approach (see, e.g., [2, 7, 11, 13, 16, 23]), namely, from the idea of describing imprecise (fuzzy) expert knowledge, of the type “if  $t$  is close to  $t_i$ , then  $x(t)$  is close to  $x(t_i)$ ”. From this viewpoint, the function  $A(t)$  is a membership function that corresponds to the word “close”.

**A somewhat unexpected empirical fact.** Intuitively, one would expect that the closer the function  $A(t)$  to how we actually think, the more successful would be the results. Empirical studies show that rather complex membership functions are needed to represent our reasoning; see, e.g., [16]. However, surprisingly, in many of these applications, many good results are obtained when researchers use a very simple triangular membership function  $A(t) = 1 - |t|$ . Why?

*Comment.* It is worth mentioning that triangular membership functions are successfully used in other applications of fuzzy techniques; for example:

- in fuzzy control; see, e.g., an application to control of telerobots in space medicine [5];
- in information accessing systems such as information retrieval systems, filtering systems, recommender systems, and web quality evaluation tools; see, e.g., [6] and references therein, etc.

Piece-wise linear functions – similar to triangular membership functions – are also effectively used in neural networks; see, e.g., [4, 21, 22].

**It is desirable to have theoretical explanations for this empirical fact.** 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 successfully, and even helps develop even more efficient techniques.

Moreover, since each theoretical justification increases our confidence in the corresponding method, the more different explanations we have, the higher our

confidence. Thus, it is better to have as many different theoretical explanations as possible.

From this viewpoint, it is desirable to have as many theoretical explanations for possible for the above empirical fact – that triangular membership functions, in spite of them being different from what we elicit from the experts, have been successfully used in many applications of F-transforms (and in other applications of fuzzy techniques).

**How this empirical fact is explained now.** At present, there are several theoretical applications for the success of triangular membership functions.

In [3], [9], and [10], it is shown that these membership functions are the most *robust* – in the sense that a given change in the input  $x$  to  $x' \approx x$  leads to the smallest possible change in the value  $A(x)$  of the corresponding membership function; specifically:

- In [10], we describe this requirement in crisp terms – as minimizing the difference between the values of  $A(x)$  and  $A(x')$ .
- In contrast, in [3] and [9], we consider this requirement in fuzzy terms – as the requirement that the degree of closeness between  $A(x)$  and  $A(x')$  should be the largest possible. In [3], this is done with type-1 fuzzy techniques, and in [9], with type-2 fuzzy techniques.

In [8], an alternative explanation was proposed: namely, it was shown that the triangular membership functions are the least vulnerable to noise.

It should be mentioned that these explanations are “local” in the sense that they consider:

- either values  $A(x)$  at nearby points  $x$
- or the effect of noise – which is also added locally, for each  $x$  separately.

**What we do in this paper.** In this paper, we provide an alternative “global” explanation for the successful use of triangular membership functions: namely, we show that the triangular membership functions are the only ones whose use enables us to correctly reconstruct an appropriate global characteristic. For this purpose:

- first, we consider the selection of an appropriate global characteristic, solve the resulting optimization problem, and thus, find the global characteristic which is optimal in some reasonable sense;
- second, we prove that the triangular membership functions are the only ones that allow us to uniquely reconstruct the optimal global characteristic of the original signal.

These two results are new. The second result is the main contributions of this paper. The first result is auxiliary for this paper – but it may be useful for other applications.

**The structure of the paper.** We want to find all membership functions that allow us to exact reconstruct the most adequate global characteristic. This idea is described in Section 2. To find the resulting membership functions, we first describe which characteristics to consider and which characteristics are the most adequate. In Section 3, we provide a general description of what we mean by a global characteristic. In Section 4, we use a general description of optimization problems to provide a natural formalization for what it means for characteristics to be more (or less) adequate than others.

The resulting optimization problem is then solved in Section 5, where we describe the global characteristics which are optimal with respect to the corresponding optimality criteria. Finally, Section 6 contains the main result of this paper: that only triangular membership functions enable us to uniquely reconstruct the most adequate global characteristic of the original signal from its F-transform.

## 2 Local Vs. Global Characteristics: Main Idea

**What we mean by local and global characteristics.** No measuring instrument can provide an instantaneous value of a physical quantity. No matter at what time  $t$  we perform our measurement, the measurement result depends not only on the value of the signal  $x(t)$  at this moment of time, but also on the values  $x(s)$  at nearby moments of time.

In some cases, we are interested in the *local* behavior of the signal. In this case, we try to measure values which are as close to  $x(t)$  as possible. F-transform values are an example of such a local analysis.

In other cases, we are interested in the *global* trend. In such cases, instead of concentrating on a short-term time interval, we deliberately measure the signal over a long period of time.

**Resulting idea.** To most adequately reconstruct the signal, we should be able to adequately reproduce both its local and its global characteristics. By definition, F-transform adequately represents the local characteristics, no matter what membership function  $A(t)$  we select. So, it is reasonable to select a membership function which most adequately represents global characteristics.

Let us describe this idea in precise terms.

## 3 Which Global Characteristics Should We Represent: Discussion

**Need for linearization.** Signals are usually weak. Thus, for any quantity  $q$  that depends on this signal  $x(t)$  – be it local or global – we should be able to ignore terms which are quadratic or higher order in terms of  $x(t)$  and thus retains only the linear terms in the corresponding dependence. As a result, we

should only consider linear quantities, i.e., quantities of the type

$$q = \int q(t) \cdot x(t) dt. \tag{3}$$

**Which linear quantities should we select?** Of course, when we perform F-transform, we lose some information about the signal. Indeed, on each time interval, we replace infinitely many values  $x(t)$  corresponding to infinitely many moments of time  $t$  from this interval, with finite many values of the corresponding F-transform. Thus, we cannot perfectly reconstruct all possible global characteristics  $q$ , since from the values of all these characteristics (e.g., of the integrals  $\int_{-\infty}^t x(s) ds$ ) we would be able to uniquely reconstruct all the values  $x(t)$ .

Thus, we need to select the most appropriate global characteristics.

**How to define what is most appropriate?** In different situations, different global characteristics may be more appropriate. In this paper, instead of trying to list specific notions of appropriateness, we will consider all possible optimality criteria of this type.

Interestingly, it turns out that all reasonable optimality criteria of this type lead, in effect, to the same family of optimal global characteristics – and the only way to reconstruct these characteristics exactly is to use triangular membership functions.

Let us describe all this in precise terms.

## 4 Selecting the Most Adequate Global Characteristic: Towards Precise Formulation of the Problem

**Towards describing what is more appropriate and what is less appropriate.** As we have mentioned, all global characteristics have the form  $q = \int q(t) \cdot x(t) dt$ . Thus, selecting a characteristic is equivalent to selecting the corresponding function  $q(t)$ .

This function  $q(t)$  may be discontinuous, as in the above example of a characteristic  $\int_{-\infty}^t x(s) ds$ . However, at least it should be measurable (non-measurable functions cannot be defined without using the Axiom of Choice, which means that they are not definable).

Of course, if we can reconstruct the value  $\int q(t) \cdot x(t) dt$ , then, for every real value  $c$ , we can also reconstruct the related value  $\int (c \cdot q(t)) \cdot x(t) dt$ , since this related value is simply equal to  $c \cdot \int q(t) \cdot x(t) dt$ . Thus, strictly speaking, a characteristics is represented not by a single function, but by the entire family  $\{c \cdot q(t)\}_{c \neq 0}$  of the related functions. So, we arrive at the following definition.

**Definition 1.** *By a characteristic or, alternatively, a family, we mean a family of the type  $\{c \cdot q(t)\}_{c \neq 0}$ , where  $q(t)$  is a given measurable function, and  $c$  runs over all possible non-zero real numbers.*

**Discussion.** To describe which families are most appropriate – i.e., to formulate the corresponding optimization problem – we need to describe, in precise terms, what it means that some characteristic (family) is more appropriate than the other one.

To come up with such a formalization, let us recall how optimization problems are (and can be) described in general. In the general case, we have a set of alternatives  $\mathcal{A}$ , and for alternatives from this set, we have defined:

- what it means for the alternative  $a$  to be better than the alternative  $b$  (we will denote it by  $a \prec b$ ), and
- what it means for the alternatives  $a$  and  $b$  to be of the same quality (we will denote it by  $a \sim b$ ).

In these general terms, an alternative  $a_{\text{opt}}$  is *optimal* if it is better than or equivalent to any other alternatives, i.e., if for every alternative  $a \in \mathcal{A}$ , we have either  $a_{\text{opt}} \prec a$  or  $a_{\text{opt}} \sim a$ .

In the traditional formulation of an optimization problem, we have an objective function  $f(a)$  defined on the set  $\mathcal{A}$  of all possible alternatives. We want to minimize the value of this objective function. (Sometimes, we want to maximize it; this case can be treated similarly.) In this case, the above relations  $a \prec b$  and  $a \sim b$  can be naturally defined in terms of this objective function, namely:

- we have  $a \prec b$  if and only if  $f(a) < f(b)$ ; and
- we have  $a \sim b$  if and only if  $f(a) = f(b)$ .

In this formulation, an alternative  $a_{\text{opt}}$  is optimal if it attains the smallest possible value of the objective function, i.e., if  $f(a_{\text{opt}}) \leq f(a)$  for all  $a \in \mathcal{A}$ .

In some cases, there are several different alternatives which are all optimal in the sense of the given criterion. For example, if the set  $\mathcal{A}$  contains algorithms for solving a certain class of problems, and  $f(a)$  is the average approximation error of the solution, we may have several different methods that guaranteed the same smallest possible average approximation error.

In such cases, we can use this uncertainty to optimize some other objective function  $g(a)$ . For example, in the algorithms case, we can select, among all the algorithms with the smallest possible average approximation error  $f(a)$ , the one for which, e.g., the worst-case approximation error is the smallest possible.

In such situations, the alternative  $a$  is better than the alternative  $b$  if:

- either  $a$  is better than  $b$  with respect to the original optimality criterion, i.e.,

$$f(a) < f(b), \tag{5}$$

- or with respect to the original optimality criterion, the alternatives  $a$  and  $b$  are of equal quality, but the alternative  $a$  is better with respect to the second objective function, i.e., if

$$f(a) = f(b) \text{ and } g(a) < g(b). \tag{6}$$

In other words, in such situations, we can have a more complex way of describing when an alternative  $a$  is better than the alternative  $b$ :

- we have  $a \prec b$  if and only if

$$f(a) < f(b) \text{ or } (f(a) = f(b) \text{ and } g(a) < g(b)); \quad (7)$$

and

- we have  $a \sim b$  if and only if

$$f(a) = f(b) \text{ and } g(a) = g(b). \quad (8)$$

Even with this modified optimality criterion, we can still have several equally optimal alternatives. For example, we may have several different algorithms which have the same smallest possible average approximation error and the same smallest possible worst-case approximation error. In such cases, we can use this non-uniqueness to optimize yet another objective function  $h(a)$  – e.g., the average computation time, etc.

How can we describe such general optimization settings in precise terms? In all these cases, what is important is which alternatives are better than others and which are of the same quality. In other words, we need some criterion according to which, for every two alternatives  $a$  and  $b$ , we can say one of the three things:

- we can say that  $a$  is better than  $b$  in the sense of this criterion; we will denote this by

$$a \prec b; \quad (9)$$

- we can say that  $b$  is better than  $a$  in the sense of the given criterion; we will denote this by

$$b \prec a; \quad (10)$$

- or we can say that the two alternatives are equally good with respect to the given criterion; we will denote this by  $a \sim b$ .

No matter what is the optimality criterion, we have these relations. Thus, we can simply make these relations the definition of an optimality criterion.

Of course, we need to make sure that these relations are consistent: e.g., if  $a$  is better than  $b$  and  $b$  is better than  $c$ , then  $a$  should be better than  $c$ . Thus, we arrive at the following definition.

**Definition 2.** (see, e.g., [12]) *Let  $\mathcal{A}$  be a set; its elements will be called alternatives. By a optimality criterion, we mean a pair of relations  $\langle \prec, \sim \rangle$  of the set  $\mathcal{A}$  that satisfies the following properties:*

- for every two alternatives  $a$  and  $b$ , we have one and only one of three options:  $a \prec b$ ,  $b \prec a$ , and  $a \sim b$ ;
- if  $a \prec b$  and  $b \prec c$ , then  $a \prec c$ ;

- if  $a \prec b$  and  $b \sim c$ , then  $a \prec c$ ;
- if  $a \sim b$  and  $b \prec c$ , then  $a \prec c$ ;
- if  $a \sim b$  and  $b \sim c$ , then  $a \sim c$ ;
- $a \sim a$ , and
- if  $a \sim b$ , then  $b \sim a$ .

In these terms, an alternative is optimal if it is better than or of the same quality as any other alternative:

**Definition 3.** We say that a characteristic  $a_{\text{opt}}$  is optimal with respect to the optimality criterion  $\langle \prec, \sim \rangle$  if for every alternative  $a \in \mathcal{A}$ , we have  $a_{\text{opt}} \prec a$  or  $a_{\text{opt}} \sim a$ .

**Discussion.** From the purely mathematical viewpoint, these conditions may be all we need; however, as we will show, from the practical viewpoint, we need to impose some additional requirement on the corresponding pair of relations.

Indeed, the whole purpose of selecting an optimality criterion is to use this optimality criterion for selecting the best alternative, i.e., the alternative which is better – according to this criterion – than any other alternative.

So, if for some pair of relations  $\langle \prec, \sim \rangle$ , there is no such optimal alternative, the corresponding “optimality criterion” is, from the practical viewpoint, completely useless. Thus, we need to require that there should be at least one optimal alternative.

What if there are several characteristics which are all the most appropriate according to the given criterion? In this case, as we have mentioned earlier, we can use this non-uniqueness to optimize something else. For example, if several characteristics are equally good in terms of accuracy with which we can predict the future behavior of the signal, then we can select among them the characteristic which is the easiest to compute. As a result, we get, in effect, a new optimality criterion, according to which  $a$  is better than  $b$  if:

- either  $a$  is better than  $b$  according to the original optimality criterion,
- or  $a$  is equivalent to  $b$  in terms of the original optimality criterion but better according to the additional optimality criterion.

If for the new optimality criterion, we still have several different optimal alternatives, we can then optimize something else, etc., until we reach a final optimality criterion for which there is exactly one optimal alternative.

**Definition 4.** We say that an optimality criterion is final if there exists exactly one alternative which is optimal with respect to this criterion.

**Need for scale-invariance.** Let us apply the above general formulation to the problem of selecting the most adequate global characteristic – or, to be more precise, to the problem of selecting the most adequate family  $F = \{c \cdot q(t)\}_{c \neq 0}$  of such characteristics.

In our problem, we deal with signals  $x(t)$  that describe how the value of a physical quantity  $x$  depends on time. We may have a starting point for the corresponding process, which provides a natural starting point for measuring time, but in general, the numerical value of time depends on what unit we use for measuring time. We can use seconds or minutes or hours – the time interval will be the same but the numerical values will change.

When we replace the original unit for measuring time with a new unit which is  $\lambda$  times smaller, then all numerical values of time are *re-scaled*, i.e., multiplied by  $\lambda$ . For example, if we go from seconds to milliseconds, all numerical values are multiplied by 1000. The function  $q(t)$  in the new units becomes  $q(\lambda \cdot t)$ .

It is reasonable to require that the relative quality of different characteristics should not change if we simply change the unit used for measuring time, without changing anything of substance. In other words, it is reasonable to require that the criterion be “scale-invariant”. Here is a precise definition.

**Definition 4.** *We say that an optimality criterion  $\langle \prec, \sim \rangle$  on the set of all possible families (in the sense of Definition 1) is scale-invariant if for every two functions  $q(t)$  and  $r(t)$  and for every  $\lambda > 0$ , the following two conditions hold:*

- if  $\{c \cdot q(t)\}_c \prec \{c \cdot r(t)\}_c$ , then  $\{c \cdot q(\lambda \cdot t)\}_c \prec \{c \cdot r(\lambda \cdot t)\}_c$ ;
- if  $\{c \cdot q(t)\}_c \sim \{c \cdot r(t)\}_c$ , then  $\{c \cdot q(\lambda \cdot t)\}_c \sim \{c \cdot r(\lambda \cdot t)\}_c$ .

## 5 Which Characteristics Are the Most Adequate: Auxiliary Result

**Discussion.** In the previous section, we argued that the most adequate global characteristic must be optimal with respect to some final scale-invariant optimality criterion. Let us describe all such characteristics.

**Proposition 1.** *For every final scale-invariant optimality criterion, each optimal characteristic has the form  $\{c \cdot x^\beta\}_c$ , for some real value  $\beta$ .*

*Comment.* Note that we did not select any specific optimality criterion. We could select an optimality criterion and find out which characteristic is better for this particular criterion. However, there are many possible optimality criteria. So, if we restrict ourselves to single optimality criteria, we would need to repeatedly solve the corresponding optimization problem for all these different optimality criteria.

Instead, Proposition 1 provides a *general* result that covers *all* (reasonable) optimality criteria. According to this result, no matter what optimality criterion we choose, the optimal characteristic always has the form  $c \cdot x^\beta$ . For different optimality criteria, we may have different  $\beta$ , but the form is always the same.

**Proof of Proposition 1.** Let us denote the scaling transformation that transforms a family  $F = \{c \cdot q(t)\}_c$  into a re-scaled family  $\{c \cdot q(\lambda \cdot t)\}_c$  by  $T_\lambda$ . In terms of this notation, scale-invariance means that:

- if  $F \prec G$ , then  $T_\lambda(F) \prec T_\lambda(G)$ ; and
- if  $F \sim G$ , then  $T_\lambda(F) \sim T_\lambda(G)$ .

Let  $\langle \prec, \sim \rangle$  be the final scale-invariant criterion. Since this criterion is final, there exists exactly one optimal characteristic  $F_{\text{opt}}$ . Let us prove that this characteristic is scale-invariant, i.e., that  $T_\lambda(F_{\text{opt}}) = F_{\text{opt}}$  for all  $\lambda > 0$ . (This proof is similar to the one given in [12].)

Indeed, since  $F_{\text{opt}}$  is optimal, it is better than or equivalent to any other characteristic. In particular, for every  $G$ , the characteristic  $F_{\text{opt}}$  is better than or equivalent to  $T_{1/\lambda}(G)$ :  $T_{1/\lambda}(G) \prec F_{\text{opt}}$  or  $T_{1/\lambda}(G) \sim F_{\text{opt}}$ . By applying scale-invariance, we conclude that either  $T_\lambda(T_{1/\lambda}(G)) \prec T_\lambda(F_{\text{opt}})$  or  $T_\lambda(T_{1/\lambda}(G)) \sim T_\lambda(F_{\text{opt}})$ . However, one can easily check that  $T_\lambda(T_{1/\lambda}(G)) = G$ .

Thus, for every characteristic  $G$ , we have either  $G \prec T_\lambda(F_{\text{opt}})$  or  $G \sim T_\lambda(F_{\text{opt}})$ . By definition of an optimal characteristic, this means that the characteristic  $T_\lambda(F_{\text{opt}})$  is optimal. However, for the final criterion, there is only one optimal characteristic, so we conclude that  $T_\lambda(F_{\text{opt}}) = F_{\text{opt}}$ . Thus, the optimal characteristic is indeed scale-invariant.

By definition, each characteristic has the form  $\{c \cdot q(t)\}_c$ . Let us denote the function  $q(t)$  corresponding to the optimal characteristic by  $q_{\text{opt}}(t)$ . The fact that the optimal family is scale-invariant means, in particular, that for every  $\lambda > 0$ , the function  $q_{\text{opt}}(\lambda \cdot t)$  – which belongs to the re-scaled family  $T_\lambda(F_{\text{opt}})$  – also belongs to the original family, i.e., has the form  $c(\lambda) \cdot q_{\text{opt}}(t)$  for some value  $c(\lambda)$ :  $q_{\text{opt}}(\lambda \cdot t) = c(\lambda) \cdot q_{\text{opt}}(t)$ . It is known that the only measurable functions satisfying this functional equation are functions of the type  $C \cdot t^\beta$ ; see, e.g., [1].

The proposition is proven.

## 6 Main Result: A New Justification of Triangular Membership Functions

Now that we know which global characteristics are the most adequate, let us now find out which membership functions can allow us to uniquely reconstruct the most adequate characteristic of a signal from its F-transform. We will start with a precise definition.

**Definition 5.** *We say that for a membership function  $A(t)$ , it is possible to always reconstruct a global characteristic  $\int q(t) \cdot x(t) dt$  if for every  $t_0$  and  $h$ , the value of this characteristic can be uniquely determined once we know all the values*

$$x_i = \int A\left(\frac{t-t_i}{h}\right) \cdot x(t) dt, \quad (11)$$

where  $t_i \stackrel{\text{def}}{=} t_0 + i \cdot h$ .

**Case of  $\beta = 0$ .** A particular case of the most adequate global characteristic is the case  $\beta = 0$ , when  $q(t) = \text{const}$  and the corresponding global characteristic

is simply the integral  $\int x(t) dt = 1$ . This characteristic can always be reconstructed from the F-transform, since we require that  $\sum_i A\left(\frac{t-t_i}{h}\right) = 1$  for all  $t$  and thus,

$$\int x(t) dt = \sum_i \int A\left(\frac{t-t_i}{h}\right) \cdot x(t) dt = \sum_i x_i. \quad (12)$$

**General case.** Thus, we should worry only about the case when  $\beta \neq 0$ . In this case, we have the following result.

**Proposition 2.** *The only membership function  $A(t)$  for which it is possible to always reconstruct a most adequate global characteristic with  $\beta \neq 0$  is the triangular membership function – it can reconstruct the characteristic  $\int t \cdot x(t) dt$  corresponding to  $\beta = 1$ .*

*Comment.* This result provides the desired global explanation of why triangular membership functions work so well in F-transform applications.

**Proof of Proposition 2.** Let us assume that for some  $\beta \neq 0$ , the membership function  $A(t)$  enables us to always uniquely reconstruct the corresponding characteristic

$$\int t^\beta \cdot x(t) dt. \quad (13)$$

Let us first consider the case when  $t_0 = 0$ ,  $h = 1$ , and the signal  $x(t)$  is equal to 0 everywhere except for the interval  $[0, 1]$ . Then, only two F-transform values are different from 0:

- the value  $x_0 = \int_0^1 A(t) \cdot x(t) dt$ , and
- the value  $x_1 = \int_0^1 A(t-1) \cdot x(t) dt$ .

The fuzzy partition requirement implies that  $A(t) + A(t-1) = 1$ , so

$$A(t-1) = 1 - A(t). \quad (14)$$

The only way to be able to always reconstruct the value  $\int_0^1 t^\beta \cdot x(t) dt$  from these two values, no matter how the signal  $x(t)$  behaves on the interval  $[0, 1]$ , is to have  $t^\beta$  equal to a linear combination of  $A(t)$  and  $A(t-1) = 1 - A(t)$ . Thus, the function  $t^\beta$  is a linear combination of  $A(t)$  and 1, and hence,  $A(t)$  is a linear combination of  $t^\beta$  and 1, i.e.,  $A(t) = a + b \cdot t^\beta$ .

For  $t = 1$ , we must have  $A(1) = 0$ , so  $a + b = 0$  and thus,  $A(t) = a \cdot (1 - t^\beta)$ . For  $t = 0$ , we must have  $A(0) = 1$ , so we have  $a = 1$  and  $A(t) = 1 - t^\beta$  for  $t \in [0, 1]$ . Correspondingly, for  $s \in [-1, 0]$ , due to  $A(t-1) = 1 - A(t)$ , we have

$$A(s) = 1 - A(s+1) = (s+1)^\beta. \quad (15)$$

Let us now consider a signal which is different from 0 only on the interval  $[1, 2]$ . For this signal, the desired global characteristic has the form

$\int_1^2 t^\beta \cdot x(t) dt$ , and the only non-zero values of F-transform are  $x_1 = \int_1^2 (1 - (t - 1)^\beta) \cdot x(t) dt$  and  $x_2 = \int_1^2 (t - 1)^\beta \cdot x(t) dt$ . Thus, the only way to exactly reconstruct the global characteristic is to have  $t^\beta$  to be a linear combination of  $1 - (t - 1)^\beta$  and  $(t - 1)^\beta$ , i.e., as a linear combination of  $(t - 1)^\beta$  and  $1$ :  $t^\beta = a \cdot (t - 1)^\beta + b$ .

Let us show that  $\beta = 1$ . For this, we need to show that cases when  $\beta > 1$  and  $\beta < 1$  are impossible.

Indeed, differentiating both sides by  $t$ , we get  $\beta \cdot t^{\beta-1} = a \cdot \beta \cdot (t - 1)^{\beta-1}$ . If  $\beta > 1$ , then for  $t = 1$ , we get  $\beta = 0$ , which contradicts the assumption that  $\beta > 1$ . If  $\beta < 1$ , then for  $t = 1$ , we get  $\beta = \infty$  – also a contradiction.

Thus,  $\beta = 1$ , so  $A(t) = 1 - |t|$ , i.e., we indeed have a triangular membership function. The proposition is proven.

*Comment.* Once we have a triangular membership function, it is easy to combine the F-transform values to get an integral of a linear function. For simplicity, assume that we start with the signal which is 0 for  $t < 0$ , and that  $h = 1$ . Then, the values  $x(t)$  corresponding to  $t \in [0, 1]$ , affect the value  $x_0$ , with the weight  $1 - t$ , and the value  $x_1$ , with weight  $t$ . If we take the difference  $x_1 - x_0$ , this difference corresponds to the weight  $2t - 1$  on  $[0, 1]$  (and the weight  $2 - x$  for  $x \in [1, 2]$ ).

We can normalize the difference  $x_1 - x_0$  to get the coefficient at  $t$  on  $[0, 1]$  to be equal to 1. For the resulting normalized linear combination  $\frac{1}{2} \cdot (x_1 - x_0)$ , on  $[0, 1]$ , we have the weight  $t - \frac{1}{2}$ , and on  $[1, 2]$ , the weight  $1 - \frac{t}{2}$ .

On the interval  $[1, 2]$ , the next F-transform value  $x_2$  corresponds to the coefficient  $t - 1$  (and 0 before that). Thus, by adding  $x_2$  with the appropriate coefficient, we can make sure that the linear combination continues to have  $t$  with coefficient 1 on the interval  $[1, 2]$  as well. For that, we need to add  $x_2$  with coefficient  $\frac{3}{2}$ . Then the resulting linear combination  $\frac{1}{2} \cdot (x_1 - x_0) + \frac{3}{2} \cdot x_2$  is equal to  $t - \frac{1}{2}$  on the whole interval  $[0, 2]$ .

On  $[2, 3]$ , this combination is equal to  $\frac{3}{2} \cdot (3 - t)$ . So, to make sure that we get a linear combination which is equal to  $t - \frac{1}{2}$  on the interval  $[2, 3]$  as well, we need to add  $x_3$  with coefficient  $\frac{5}{2}$ , etc. At the end, when we reach the end of the time interval on which the signal is defined, the corresponding linear combination gives us the integral

$$\int \left(t - \frac{1}{2}\right) \cdot x(t) dt = \int t \cdot x(t) dt - \frac{1}{2} \cdot \int x(t) dt. \quad (16)$$

Since, as we have mentioned, we can easily determine the integral  $\int x(t) dt$  by adding all the values of the F-transform, we can thus indeed determine the value of the desired global characteristic  $\int t \cdot x(t) dt$ .

## Author Contributions

The authors contributed equally to this work.

## Funding

This work was supported by the Banking University of Ho Chi Minh City, Vietnam, 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] Aczél, J.; Dhombres, J. *Functional Equations in Several Variables*, Cambridge University Press: Cambridge, UK, 2008.
- [2] Belohlavek, R.; Dauben, J.W.; Klir, G.J. *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press: New York, 2017.
- [3] Gholamy, A.; Kosheleva, O.; Kreinovich, V. How to explain the efficiency of triangular and trapezoid membership functions in applications to design, *Ontology of Designing* **2019**, 9(2).
- [4] Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*, MIT Press: Cambridge, Massachusetts, 2016.
- [5] Haidegger, T.; Kovács, L.; Precup, R.-E.; Benyó, B.; Benyó, Z.; Preitl, S. Simulation and control in telerobots in space medicine, *Acta Astronautica* **2012**, 81, 390–402.
- [6] Herrera-Viedma, E.; López-Herrera, A.G. A review on information accessing systems based on fuzzy linguistic modeling, *International Journal on Computational Intelligence Systems* **2010**, 3(4), 420–437.
- [7] Klir, G.; Yuan, B. *Fuzzy Sets and Fuzzy Logic*, Prentice Hall: Upper Saddle River, New Jersey, 1995.

- [8] Kosheleva, O.; Kreinovich, V. Why triangular membership functions are often efficient in F-transform applications: relation to probabilistic and interval uncertainty and to Haar wavelets”, In: *Proceedings of the 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'2018*, Medina, J. et al. (eds.), Cadiz, Spain, June 11–15, 2018.
- [9] Kosheleva, O.; Kreinovich, V.; Shahbazova, S.N. Type-2 fuzzy analysis explains ubiquity of triangular and trapezoid membership functions, In: *Proceedings of the World Conference on Soft Computing*, Shahbazova, S.N.; Kacprzyk, J.; Balas, V.E.; Kreinovich, V. (eds.), Baku, Azerbaijan, May 29–31, 2018.
- [10] Kreinovich, V.; Kosheleva, O.; Shahbazova, S.N. Why triangular and trapezoid membership functions: a simple explanation, In: *Proceedings of the World Conference on Soft Computing*, Shahbazova, S.N.; Kacprzyk, J.; Balas, V.E.; Kreinovich, V. (eds.), Baku, Azerbaijan, May 29–31, 2018.
- [11] Mendel, J.M. *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer: Cham, Switzerland, 2017.
- [12] Nguyen, H.T.; Kreinovich, V. *Applications of Continuous Mathematics to Computer Science*, Kluwer: Dordrecht, 1997.
- [13] Nguyen, H.T.; Walker, C.; Walker, E.A. *A First Course in Fuzzy Logic*, Chapman and Hall/CRC: Boca Raton, Florida, 2019.
- [14] Novak, V.; Perfilieva, I.; Holcapek, M.; Kreinovich, V. Filtering out high frequencies in time series using F-transform, *Information Sciences* **2014**, 274, 192–209.
- [15] 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*, Inuiguchi, M.; Kusunoki, Y.; Seki, M. (eds.), Osaka, Japan, September 24–27, 2012.
- [16] Novák, V.; Perfilieva, I.; Močkoř, J. *Mathematical Principles of Fuzzy Logic*, Kluwer: Boston, Dordrecht, 1999.
- [17] Perfilieva, I. Fuzzy transforms: Theory and applications, *Fuzzy Sets and Systems* **2006**, 157, 993–1023.
- [18] Perfilieva, I. F-transform, In: *Springer Handbook of Computational Intelligence*, Springer Verlag, 2015, pp. 113–130.
- [19] Perfilieva, I., Danková, M., Bede, B. Towards a higher degree F-transform”, *Fuzzy Sets and Systems* **2011**, 180(1), 3–19.

- [20] 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*, Inuiguchi, M.; Kusunoki, Y.; Seki, M. (eds.), Osaka, Japan, September 24–27, 2012.
- [21] Ruiz-Rangel, J.; Ardila Hernandez, C.J.; Maradei Gonzalez, L.; Jabba Molinares, D. ERNEAD: training of artificial neural networks based on a genetic algorithm and finite automata theory, *International Journal of Artificial Intelligence* **2018**, 16(1), 214–253.
- [22] Saadat, J.; Moallem, P.; Koofgar, H. Training echo estate neural network using harmony search algorithm, *International Journal of Artificial Intelligence* **2017**, 15(1), 163–179.
- [23] Zadeh, L.A. Fuzzy sets, *Information and Control* **1965**, 8, 338–353.