12-2017

Why Triangular Membership Functions Are Often Efficient in F-Transform Applications: Relation to Interval Uncertainty\\ and Haar Wavelets

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-17-93

Why Triangular Membership Functions Are Often Efficient in F-Transform Applications: Relation to Interval Uncertainty and Haar Wavelets

Olga Kosheleva[0000−0003−2587−4209] and
Vladik Kreinovich[0000−0002−1244−1650]

University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract. Fuzzy techniques describe expert opinions. At first glance, we would therefore expect that the more accurately the corresponding membership functions describe the expert's opinions, the better the corresponding results. In practice, however, contrary to these expectations, the simplest – and not very accurate – triangular membership functions often work the best. In this paper, on the example of the use of membership functions in F-transform techniques, we provide a possible theoretical explanation for this surprising empirical phenomenon.

1 Formulation of the Problem

Practical problem: need to find trends in observations. In many practical situations, we analyze how a certain quantity x changes with time t. For example, we may want to analyze how an economic characteristic changes with time:

- we want to analyze the trends,
- we want to know what caused these trends, and
- we want to make predictions and recommendations based on this analysis.

To perform this analysis, we observe the values $x(t)$ of the desired quantity at different moments of time t. Often, however, the observed values themselves do not provide a good picture of the corresponding trends, since the observed values contain some random (noise-type) factors that prevent us from clearly seeing the trends.

For economic characteristics such as the stock market value, on top of the trend – in which we are interested – there are always day-by-day and even hour-by-hour fluctuations. For physical measurements, a similar effect can be caused by measurement uncertainty, as a result of which the measured values $x(t)$ differ from the clear trend by a random measurement error – error that differs from one measurement to another.

How can we detect the desired trend in the presence of such random noise?

F-transform approach to solving this problem: a brief reminder. One of the successful approach for solving the above trend-finding problem comes from the F-transform idea; see, e.g., [13, 14, 16–19].

One of the ideas behind F-transform comes from the fact that what we really want is not just a *quantitative* mathematical model, we want a good *qualitative* understanding of the corresponding trend – and of how this trend changes with time. For example, we want to be able to say that the stock market first somewhat decreases, then rapidly increases, etc. In other words, we want these trends to be described in terms of time-localized natural-language properties.

Once we selected these properties, we can use fuzzy logic techniques (see, e.g., [1, 9, 6, 12, 15, 22]) to describe these properties in computer-understandable terms, as time-localized membership functions $x_1(t)$, ..., $x_n(t)$. Time-localized means that when we analyze the process $x(t)$ on a wide time interval $[\underline{T}, \overline{T}]$:

- the first membership function $x_1(t)$ is different from 0 only on a narrow interval $[\underline{T}_1, \overline{T}_1]$, where $\underline{T}_1 = \underline{T}$;
- the second membership function $x_2(t)$ is different from 0 only on a narrow interval $[\underline{T}_2, \overline{T}_2]$, where $\underline{T}_2 \le \overline{T}_1$;
- etc.

so that the whole range $[\underline{T}, \overline{T}]$ is covered by the corresponding ranges $[\underline{T}_i, \overline{T}_i]$.

Once we have these functions $x_i(t)$, then, as a good representation of the original signal's trend, it is reasonable to consider, e.g., linear combinations

$$x_a(t) = \sum_{i=1}^{n} c_i \cdot x_i(t) \tag{1}$$

of these functions as the desired reconstruction for the no-noise signal.

This approach has indeed led to many successful applications.

In many practical applications, triangular membership functions work well. Which membership functions should we use in this approach? At first glance, since the objective of a membership function is to capture the expert reasoning, we may expect that the more adequately these functions capture the expert reasoning, the more adequate will be our result. From this viewpoint, we expect complex membership functions to work the best.

Somewhat surprisingly, however, in many practical applications, the simplest possible triangular membership functions work the best, i.e., functions of the type

$$x_i(t) = \max\left(1 - \frac{|x - c|}{w}, 0\right)$$

that:

- linearly rise from 0 to 1 on the interval $[c - w, c]$, and then
- linearly decrease from 1 to 0 on the interval $[c, c + w]$.

Why? The above empirical fact needs explanation: why triangular membership functions work so well?

What we do in this paper. In this paper, we provide a possible explanation for this empirical phenomenon.

2 Analysis of the Problem and the Main Ideas Behind Our Explanation

What is a trend: discussion. As we have mentioned earlier, we are interested *not* so much in predicting the moment-by-moment values of the corresponding quantity $x(t)$ – these values contains random fluctuations. What we are interested in is the *trend*. So, to analyze this problem in precise terms, we need to understand what we mean by a trend.

A trend may mean increasing or decreasing, decreasing fast vs. decreasing slow, etc. In the ideal situation, in which we do not have any random fluctuations, all these properties can be easily described in terms of the time derivative $x'(t) \stackrel{\text{def}}{=} \dfrac{dx}{dt}$ of the corresponding process.

From this viewpoint, understanding the trend means reconstructing the *derivative* $x'(t)$ of the observed process based on its random-fluctuation-corrupted observed values.

What is F-transform from this viewpoint. We are interested in the trend, so once we have applied the F-transform technique and obtained the desired no-boise expression (1), what we really want is to use its derivative

$$x'_a(t) = \sum_{i=1}^{n} c_i \cdot x'_i(t). \tag{2}$$

If we denote the derivatives $x'_i(t)$ of the membership functions by $e_i(t)$, the formula (2) then means that we approximate the derivative $e(t) \stackrel{\text{def}}{=} x'(t)$ of the original signal by a linear combination of the functions $e_i(t)$:

$$e(t) \approx e_a(t) = \sum_{i=1}^{n} c_i \cdot e_i(t). \tag{3}$$

In these terms, we approximate the original derivative by a function from a linear space spanned by the functions $e_i(t)$. In this sense, selecting the functions $x_i(t)$ means selecting the proper linear space – i.e., the proper functions $e_i(t)$.

For computational convenience, it makes sense to select an orthonormal basis. What is important is the linear space.

Each linear space can have many possible bases. From the computational viewpoint, it is often convenient to use orthonormal bases, i.e., bases for which:

- we have $\int e_i^2(t)\,dt = 1$ for all i, and

– we have $\int e_i(t) \cdot e_j(t)\,dt = 0$ for all $i \neq j$.

Thus, without losing generality, we can assume that the basis $e_i(t)$ is orthonormal.

Comment. For the typically used equally spaced triangular functions on intervals $[\underline{T}_i, \overline{T}_i] = [\underline{T} + (i-1) \cdot h, \underline{T} + (i+1) \cdot h]$, for some $h > 0$, the corresponding derivatives $e_i(t)$ are indeed orthogonal, i.e., we indeed have $\int e_i(t) \cdot e_j(t)\,dt = 0$ for all $i \neq j$, but, in general, we have

$$\int e_i^2(t)\,dt = 2h \cdot \left(\frac{1}{h}\right)^2 = \frac{1}{2h} \neq 1.$$

However, it is easy to transform this basis into an orthonormal one without changing the corresponding linear space: namely, it is sufficient to consider the new functions $e_i^*(t) = \sqrt{2h} \cdot e_i(t)$.

Mathematical analysis of the problem. Once we know the original function $e_a(t)$ and we have selected the basis $e_i(t)$, what are the parameters c_i that provide the best approximation?

We start with a tuple $e \stackrel{\text{def}}{=} (e(t_1), e(t_2), \ldots)$ that contains all observations – to be more precise, numerical derivatives $e(t_k) = \dfrac{x(t_{k+1}) - x(t_k)}{t_{k+1} - t_k}$ based on these observations. Once we have an approximating function $e_a(t)$, we can form a similar tuple based on the approximating values: $e_a \stackrel{\text{def}}{=} (e_a(t_1), e_a(t_2), \ldots)$ It is reasonable to select the coefficients c_i for which the new tuple is the closest to the original one, i.e., for which the distance

$$\sqrt{(e_a(t_1) - e(t_1))^2 + (e_a(t_2) - e(t_2))^2 + \ldots}$$

between the tuples e_a and e is the smallest possible. Since the square $z \to z^2$ is a monotonic function, minimizing the distance is equivalent to minimize the square of the distance, i.e., the quantity

$$(e_a(t_1) - e(t_1))^2 + (e_a(t_2) - e(t_2))^2 + \ldots$$

In most practical situations, measurements are performed at regular intervals, so this sum is proportional to the corresponding integral

$$\int (e_a(t) - e(t))^2\,dt.$$

So, we want to find the values c_i for which this integral attains its smallest possible value. Since we assumed that the basis is orthonormal, the optimal coefficients c_i can be simply obtained as

$$c_i = \int e(s) \cdot e_i(s)\,ds. \tag{4}$$

Thus, the representation (3) takes the form

$$e(t) \approx e_a(t) = \sum_{i=1}^{n} e_i(t) \cdot \left(\int e(s) \cdot e_i(s)\, ds \right). \tag{5}$$

We want to select the functions $e_i(t)$ for which the noise has the least effect on the result. The whole purpose of this analysis is to eliminate the noise – or at least to decrease its effect. From this viewpoint, it is reasonable to select the functions $e_i(t)$ for which the effect of the noise on the reconstructed signal $e_a(t)$ is as small as possible.

According to the formula (5), the function $e_a(t)$ is the sum of n values

$$v_i(t) \stackrel{\text{def}}{=} e_i(t) \cdot \left(\int e(s) \cdot e_i(s) \right) ds. \tag{6}$$

Thus, it is desirable to make sure that the effect of noise on each of these values v_i is as small as possible.

Noise $n(t)$ means that instead of the original function $e(t)$, we have a noise-infected function $e(t) + n(t)$. If we use this noisy function instead of the original function $e(t)$, then, instead of the original value $v_i(t)$, we get a new value

$$v_i^{\text{new}}(t) = e_i(t) \cdot \left(\int (e(s) + n(s)) \cdot e_i(s)\, ds \right). \tag{7}$$

The difference $\Delta v_i(t) = v_i^{\text{new}}(t) - v_i(t)$ between the new and the original values is thus equal to

$$\Delta v_i(t) = e_i(t) \cdot \left(\int n(s) \cdot e_i(s)\, ds \right). \tag{8}$$

This difference depends on time t and on the noise $n(t)$. To make sure that we reconstruct the trend correctly, it makes sense to require that for all possible moments of time t and for all possible noises $n(t)$, this difference does not exceed a certain value – and this value should be as small as possible. In other words, we would like to minimize the worst-case value of this difference:

$$J \stackrel{\text{def}}{=} \max_{t, n(t)} \left| e_i(t) \cdot \left(\int n(s) \cdot e_i(s)\, ds \right) \right|. \tag{9}$$

What noises $n(t)$ should we consider? In principle, in different situations, we can have different types of noise, with different statistical characteristics. What they all have in common is that usually, there is an upper bound Δ on the value of the noise: $|n(t)| \le \Delta$; see, e.g., [5, 8, 11, 20]. (In this case, $e(t) + n(t) \in [e(t) - \Delta, e(t) + \Delta]$, i.e., we have an *interval uncertainty*.)

So, we arrive at the following mathematical problem.

3 Selecting the Best Functions: Precise Formulation of the Problem and Its Solution

Definition 1. *Let us assume that we are given:*

- *the value* $\Delta > 0$*, and*
- *an interval* $[\underline{T}_i, \overline{T}_i]$.

We consider functions $e_i(t)$ *defined on the given interval for which* $\int e_i^2(t) = 1$*. For each such function* $e_i(t)$*, we define its* degree of noise-dependence *as the value*

$$J(e_i) = \max_{t,n(t)} \left| e_i(t) \cdot \left(\int n(s) \cdot e_i(s)\, ds \right) \right|, \tag{10}$$

where the maximum is taken:

- *over all moments of time* $t \in [\underline{T}_i, \overline{T}_i]$*, and*
- *over all functions* $n(t)$ *for which* $|n(t)| \le \Delta$ *for all* t.

We say that the function $e_i(t)$ *is* optimal *if its degree of noise-dependence is the smallest possible.*

Proposition 1. *A function* $e_i(t)$ *is optimal if and only if* $|e_i(t)| = \text{const}$ *for all* t.

Discussion. We usually consider membership functions $x_i(t)$ which:

- first increase, and
- then decrease.

For such functions $x_i(t)$, the derivative $e_i(t) = x_i'(t)$ is:

- first positive, and
- then negative.

Thus, for the optimal function, we:

- first have $e_i(t)$ equal to a positive constant c, and
- then equal to minus this same constant.

By integrating this piece-wise constant function, we conclude that the function $x_i(t)$:

- first linearly increases,
- then linearly decreases with the same slope,

i.e., that $x_i(t)$ is a triangular membership function.

Thus, *we have indeed explained why triangular membership functions are often efficient in F-transform applications.*

Comment. The piece-wise constant functions described above are well-known: they are known as *Haar wavelets*; see, e.g., [3, 7, 10, 21]. These functions indeed

form a basis, and often, by using this basis to approximate signals and images, practitioners get very good results.

From this viewpoint, the use of triangular membership functions in F-transform techniques is equivalent to using Haar wavelets to approximate the corresponding trend. Since Haar wavelets are known to be practically efficient, it is not surprising that F-transform techniques using triangular membership functions are practically efficient as well.

Proof of Proposition 1. The desired objective function J is the largest value of the quantity

$$q(t, n(t)) \stackrel{\text{def}}{=} \left| e_i(t) \cdot \left(\int n(s) \cdot e_i(s)\, ds \right) \right| = |e_i(t)| \cdot \left| \int n(s) \cdot e_i(s)\, ds \right| \tag{12}$$

over all possible values of t and $n(t)$:

$$J = \max_{t, n(y))} q(t, n(t)). \tag{13}$$

This double maximum can be equivalently described as

$$J = \max_{n(t)} Q(n(t)), \tag{14}$$

where we denoted

$$Q(n(t)) \stackrel{\text{def}}{=} \max_t q(t, n(t)). \tag{15}$$

Once the noise function $n(t)$ is fixed, the value

$$q(t, n(t)) \tag{16}$$

is proportional to $|e_i(t)|$. Thus, the maximum of $q(t, n(t))$ over t is attained when $|e_i(t)|$ is the largest:

$$Q(n(t)) = \max_t q(t, n(t)) = \left(\max_t |e_i(t)| \right) \cdot \left| \int n(s) \cdot e_i(s)\, ds \right|, \tag{17}$$

i.e.,

$$Q(n(t)) = \left(\max_t |e_i(t)| \right) \cdot F(n(t)), \tag{18}$$

where we denoted

$$F(n(t)) \stackrel{\text{def}}{=} \left| \int n(s) \cdot e_i(s)\, ds \right|. \tag{19}$$

The first factor in the formula (18) is a positive constant not depending on the noise $n(t)$. So, to find the largest value of $Q(n(t))$, we need to find the largest possible value of $F(n(t))$:

$$J = \max_{n(t)} Q(n(t)) = \left(\max_t |e_i(t)| \right) \cdot \max_{n(t)} F(n(t)). \tag{20}$$

The absolute value of the sum does not exceed the sum of absolute values, so

$$F(n(t)) = \left|\int n(s)\cdot e_i(s)\,ds\right| \le \int |n(s)\cdot e_i(s)|\,ds = \int |n(s)|\cdot|e_i(s)|\,ds. \quad (21)$$

For each s, we have $|n(s)| \le \Delta$, hence

$$F(n(t)) \le \Delta \cdot \int |e_i(s)|\,ds. \quad (22)$$

On the other hand, for $n(s) = \Delta\cdot\operatorname{sign}(e_i(s))$, we have

$$n(s)\cdot e_i(s) = \Delta\cdot\operatorname{sign}(e_i(s))\cdot e_i(s) = \Delta\cdot|e_i(s)|. \quad (23)$$

Hence, for this particular noise, we have

$$F(n(t)) = \left|\int \Delta\cdot|e_i(s)|\,ds\right| = \Delta\cdot\int |e_i(s)|\,ds. \quad (24)$$

So, the upper bound in the inequality (22) is always attained, hence

$$\max_{n(t)} F(n(t)) = \Delta\cdot\int |e_i(s)|\,ds. \quad (25)$$

Substituting the expression (25) into the formula (20), we conclude that

$$J = \left(\max_t |e_i(t)|\right)\cdot\Delta\cdot\int |e_i(s)|\,ds. \quad (26)$$

We want to find a function $e_i(t)$ for which this expression is the smallest possible. To find this $e_i(t)$, it is convenient to take into account that both e_i-dependent factors in the formula (26) correspond to known norms of the function $e_i(t)$ (see, e.g., [4]):

- the expression $\max\limits_t |e_i(t)|$ is the L^∞-norm $\|e_i\|_{L^\infty}$, and
- the expression $\int |e_i(s)|\,ds$ is the L_1-norm $\|e_i\|_{L^1}$.

Thus, we have

$$J = \Delta\cdot\|e_i\|_{L^\infty}\cdot\|e_i\|_{L^1}. \quad (27)$$

We consider the functions $e_i(t)$ for which $\int e_i^2(t)\,dt = 1$. This property can also be described in terms of a standard norm: namely, it can be described as $\|e_i\|_{L^2} = 1$, where

$$\|e_i(t)\|_{L^2} \stackrel{\text{def}}{=} \sqrt{\int e_i^2(t)\,dt}. \quad (28)$$

There is a known inequality connecting these three norms: Hölder's inequality (see, e.g., [4]):

$$\|f\|^2_{L^2} \le \|f\|_{L^1}\cdot\|f\|_{L^\infty}, \quad (29)$$

for which it is known that the equality is attained if and only if $|f(t)|$ is constant – wherever it is different from 0.

In our case, this inequality implies that

$$J = \Delta \cdot \|e_i\|_{L^\infty} \cdot \|e_i\|_{L^1} \geq \Delta \cdot \|e_i\|_{L^2}^2 = \Delta \cdot 1 = \Delta, \tag{30}$$

and that the smallest possible value Δ is attained when $|e_i(x)|$ is constant. This is exactly what we wanted to prove.

Comment. It should be mentioned that the ideas of this proof are similar to the ideas from our paper [2].

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. A. E. Brito and O. Kosheleva, "Interval + image = wavelet: for image processing under interval uncertainty, wavelets are optimal", *Reliable Computing*, 1998, Vol. 4, No. 3, pp. 291–301.
3. C. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, California, 1992.
4. R. E. Edwards, *Functional Analysis: Theory and Applications*, Dover, New York, 2011.
5. L. Jaulin, M. Kiefer, O. Dicrit, and E. Walter, *Applied Interval Analysis*, Springer, London, 2001.
6. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
7. S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Burlington, Massachusetts, 2008.
8. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
9. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
10. Y. Meyer, *Wavelets Algorithms and Applications*, SIAM, Philadelphia, 1993.
11. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
12. H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
13. V. Novak, I. Perfilieva, M. Holcapek, and V. Kreinovich, "Filtering out high frequencies in time series using F-transform", *Information Sciences*, 2014, Vol. 274, pp. 192–209.
14. V. Novak, I. Perfilieva, and V. Kreinovich, "F-transform in the analysis of periodic signals", In: M. Inuiguchi, Y. Kusunoki, and M. Seki (eds.), *Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012*, Osaka, Japan, September 24–27, 2012.

15. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
16. I. Perfilieva, "Fuzzy transforms: Theory and applications", *Fuzzy Sets and Systems*, 2006, Vol. 157, pp. 993–1023.
17. I. Perfilieva, "F-transform", In: *Springer Handbook of Computational Intelligence*, Springer Verlag, 2015, pp. 113–130.
18. I. Perfilieva, M. Danková, and B. Bede, "Towards a higher degree F-transform", *Fuzzy Sets and Systems*, 2011, Vol. 180, No. 1, pp. 3–19.
19. I. Perfilieva, V. Kreinovich, and V. Novak, "F-transform in view of trend extraction", In: M. Inuiguchi, Y. Kusunoki, and M. Seki (eds.), *Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012*, Osaka, Japan, September 24–27, 2012.
20. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.
21. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, New Jersey, 1995.
22. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.