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

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Practical Problem: Need to Find Trends

- 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.
- Indeed, the observed values contain some random factors that prevent us from clearly seeing the trends.

2. Need to Find Trends (cont-d)

- For economic characteristics such as the stock market:
  - 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, the measured values $x(t)$ differ from the clear trend by a random measurement error.
- This error differs from one measurement to another.
- How can we detect the desired trend in the presence of such random noise?

3. 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.
- We want not only a *quantitative* mathematical model.
- We want a good *qualitative* understanding of the corresponding trend – and of how it changes with time.
- For example, we want to be able to say that the stock market first somewhat decreases, then rapidly increases.
- In other words, we want these trends to be described in terms of time-localized natural-language properties.
- First, we select these properties.
- Then, we can use fuzzy logic techniques to describe these properties in computer-understandable terms.

4. F-Transform Approach (cont-d)

- So, we get time-localized membership functions

$$x_1(t), \ldots, x_n(t).$$

- Time-localized means that when we analyze the process $x(t)$ on a wide time interval $[\underline{T}, \overline{T}]$:
  - the 1st 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 2nd membership function $x_2(t)$ is $\neq 0$ only on a narrow interval $[\underline{T}_2, \overline{T}_2]$, where $\underline{T}_2 \leq \overline{T}_1$, etc.
- The whole range $[\underline{T}, \overline{T}]$ is covered by the corresponding ranges $[\underline{T}_i, \overline{T}_i]$.

5. F-Transform Approach (cont-d)

- 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 $x_a(t) = \sum_{i=1}^{n} c_i \cdot x_i(t)$ of these functions;
  - this will be the desired reconstruction for the no-noise signal.
- This approach has indeed led to many successful applications.

6. In Many Practical Applications, Triangular Membership Functions Work Well

- Which membership functions should we use in this approach?
- The objective of a membership function is to capture the expert reasoning.
- So, 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.
- However, in many practical applications, the simplest possible triangular membership functions work the best:

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

7. Triangular Functions: Why?

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

- These functions:
  - linearly rise from 0 to 1 on the interval $[c - w, c]$, and then
  - linearly decrease from 1 to 0 on $[c, c + w]$.
- The above empirical fact needs explanation: why triangular membership functions work so well?
- In this talk, we provide a possible explanation for this empirical phenomenon.

8. What Is a Trend: Discussion

- A trend may mean increasing or decreasing, decreasing fast vs. decreasing slow, etc.;
  - in the ideal situation with no random fluctuations,
  - all these properties can be easily described in terms of the time derivative $x'(t) \stackrel{\text{def}}{=} \dfrac{dx}{dt}$.
- From this viewpoint, understanding the trend means reconstructing the *derivative* $x'(t)$; so:
  - once we have applied the F-transform technique and obtained the desired no-noise expression

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

  - what we really want is to use its derivative

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

9. What Is a Trend: Discussion (cont-d)

- So, we must:
  - approximate the derivative $e(t) \stackrel{\text{def}}{=} x'(t)$ of the original signal
  - by a linear combination of the derivatives $e_i(t) \stackrel{\text{def}}{=} x'_i(t)$:

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

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

10. 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.
- Typically, we use used equally spaced triangular functions on intervals $[\underline{T}_i, \overline{T}_i] = [\underline{T}+(i-1)\cdot h, \underline{T}+(i+1)\cdot h]$.
- The corresponding derivatives $e_i(t)$ are indeed orthogonal.

11. Orthonormal Basis (cont-d)

- In general, $\int e_i^2(t)\,dt = 2h \cdot \left(\frac{1}{h}\right)^2 = \frac{2}{h} \neq 1.$
- However, it is easy to transform this basis into an orthonormal one: take $e_i^*(t) = \sqrt{\frac{h}{2}} \cdot e_i(t)$.
- Once we know the original function $e_a(t)$ and we have selected the basis $e_i(t)$, what are the parameters c_i?
- We start with a tuple $e \stackrel{\text{def}}{=} (e(t_1), e(t_2), \ldots)$, where $e(t_k) = \frac{x(t_{k+1}) - x(t_k)}{t_{k+1} - t_k}$.
- Once we have an approximating function $e_a(t)$, we can form a similar tuple $e_a \stackrel{\text{def}}{=} (e_a(t_1), e_a(t_2), \ldots)$
- It is reasonable to select c_i for which the distance between e_a and e is the smallest:

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

12. Orthonormal Basis (cont-d)

- This is equivalent to minimizing

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

- In most practical situations, measurements are performed at regular intervals.
- So this sum is proportional to the integral

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

- We want to find c_i for which this integral attains its smallest value; then, $c_i = \int e(s) \cdot e_i(s)\, ds$, hence:

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

13. 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.
- So, we should select $e_i(t)$ for which the effect of the noise on the reconstructed signal $e_a(t)$ is the smallest.
- $e_a(t)$ is the sum of n values $v_i(t) \stackrel{\text{def}}{=} e_i(t) \cdot \left(\int e(s) \cdot e_i(s)\right)\, ds$.
- 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 $v_i(t)$, we get:

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

14. How to Select the Functions $e_i(t)$ (cont-d)

- Reminder: $v_i(t) \stackrel{\text{def}}{=} e_i(t) \cdot \left(\int e(s) \cdot e_i(s)\right)\, ds$ and

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

- 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

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

15. What Noises $n(t)$ Should We Consider?

- In different situations, we can have different types of noise, with different statistical characteristics.
- In some cases, we know the probability distribution of the noise, i.e., we have *probabilistic uncertainty*.
- In other cases, we do now know the probabilities of different noise values.
- The only information that we have is an upper bound Δ on the value of the noise: $|n(t)| \le \Delta$.
- In this case, $e(t) + n(t) \in [e(t) - \Delta, e(t) + \Delta]$, i.e., we have an *interval uncertainty*.
- We show that in both cases, the optimal membership functions $x_i(t)$ are triangular.

16. Case of Interval Uncertainty

- The difference $\Delta v_i(t)$ 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:

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

- So, we arrive at the following mathematical problem.

17. Interval Uncertainty (cont-d)

- We are given a 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

$$J_{\rm int}(e_i) = \max_{t,n(t)} \left| e_i(t) \cdot \left(\int n(s) \cdot e_i(s)\,ds \right) \right|.$$

- Here, 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.*

18. Interval Uncertainty (cont-d)

- 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 $x_i(t)$ is triangular.
- Thus, we explained why triangular membership functions are often efficient in F-transform applications.

19. Relation to Haar Wavelets

- The piece-wise constant functions described above are known as *Haar wavelets*; so:
  - the use of triangular membership functions in F-transform techniques is equivalent to
  - using Haar wavelets to approximate the corresponding trend.
- Haar wavelets are known to be practically efficient.
- So, it is not surprising that techniques using triangular functions are practically efficient.

20. Case of Probabilistic Uncertainty

- We consider the case when for each moment t, we know the probability distribution of the noise $n(t)$.
- We do not have any reason to assume that the characteristics of noise change with time.
- So, it makes sense to assume that the variables $n(t)$ corr. to different t are identically distributed.
- We do not have any reason to assume that positive noise values are more probable than negative ones.
- So, it makes sense to assume that the distribution is symmetric, and that, as a result, its mean value is 0.
- We do not have any reason to assume that $n(t)$ and $n(t')$ are correlated.
- So, it makes sense to assume that these noises are independent, i.e., that we have a *white noise*.

21. Case of Probabilistic Uncertainty (cont-d)

- So, the difference $\Delta v_i(t)$ is a linear combination of the large number of independent variables $n_i(s)$.
- Thus, due to the Central Limit Theorem, we can conclude that the difference $\Delta v_i(t)$ is normally distributed.
- A normal distribution is uniquely determined by its mean and variance.
- Since the mean value of each $n_i(s)$ is 0, the mean of $\Delta v_i(t)$ is also 0.
- The variance of the sum of independence random variables is equal to the sum of the variances:

$$\sigma_i^2(t) = e_i^2(t) \cdot \sigma^2 \cdot \int e_i^2(s)\,ds.$$

- Here, σ characterizes the standard deviation of each noise value $n(s)$.

22. Case of Probabilistic Uncertainty (cont-d)

- Since $e_i(t)$ are orthonormal, $\int e_i^2(s)\,ds = 1$ hence

$$\sigma_i^2(t) = \sigma^2 \cdot e_i^2(t).$$

- This variance depends on the time t.
- Similarly to the interval case, it is reasonable to minimize the worst-case value $\max\limits_t (\sigma^2 \cdot e_i^2(t))$.
- Since σ^2 is a constant, minimizing this value is equivalent to minimizing the quantity $\max\limits_t e_i^2(t)$.
- So, we arrive at the following mathematical problem.

23. Case of Probabilistic Uncertainty (cont-d)

- We are given 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 $J_{\rm prob}(e_i) = \max\limits_t e_i^2(t)$.
- We say that the function $e_i(t)$ is *optimal* if its degree of noise-dependence is the smallest possible.
- **Proposition 2.** *A function $e_i(t)$ is optimal if and only if $|e_i(t)| = \text{const}$ for all t.*
- We have already shown that this implies that the original membership function $x_i(t)$ is triangular.

24. Acknowledgments

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

25. Proof of Proposition 1

- Our objective function is $J_{\rm int} = \max\limits_{t,n(y))} q(t,n(t))$, where

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

$$|e_i(t)|\cdot\left|\int n(s)\cdot e_i(s)\,ds\right|.$$

- This can be equivalently described as $J_{\rm int} = \max\limits_{n(t)} Q(n(t))$, where $Q(n(t)) \stackrel{\rm def}{=} \max\limits_t q(t,n(t))$.
- Once $n(t)$ is fixed, $q(t,n(t))$ is proportional to $|e_i(t)|$.
- Thus, $\max\limits_t q(t,n(t))$ is attained when $\max\limits_t |e_i(t)|$:

$$Q(n(t)) = \max_t q(t,n(t)) = \left(\max_t |e_i(t)|\right)\cdot F(n(t)),\ \text{ where}$$

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

26. Proof of Proposition 1 (cont-d)

- Reminder: $Q(n(t)) = \left(\max\limits_t |e_i(t)|\right) \cdot F(n(t))$.
- The first factor in this formula 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 $F(n(t))$:

$$J_{\text{int}} = \max_{n(t)} Q(n(t)) = \left(\max_t |e_i(t)|\right) \cdot \max_{n(t)} F(n(t)).$$

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

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

$$\int |n(s)| \cdot |e_i(s)|\, ds.$$

- For each s, $|n(s)| \le \Delta$, hence $F(n(t)) \le \Delta \cdot \int |e_i(s)|\, ds$.

27. Proof of Proposition 1 (cont-d)

- Reminder: $F(n(t)) \le \Delta \cdot \int |e_i(s)|\,ds$.
- On the other hand, for $n(s) = \Delta \cdot \text{sign}(e_i(s))$, we have
$$n(s) \cdot e_i(s) = \Delta \cdot \text{sign}(e_i(s)) \cdot e_i(s) = \Delta \cdot |e_i(s)|.$$
- Hence, for this particular noise, we have
$$F(n(t)) = \left| \int \Delta \cdot |e_i(s)|\,ds \right| = \Delta \cdot \int |e_i(s)|\,ds.$$
- So, the upper bound in the above inequality is always attained: $\max\limits_{n(t)} F(n(t)) = \Delta \cdot \int |e_i(s)|\,ds$.
- Substituting the expression into the formula for J_{int}, we get $J_{\text{int}} = \left(\max\limits_t |e_i(t)|\right) \cdot \Delta \cdot \int |e_i(s)|\,ds$.
- We want to find a function $e_i(t)$ for which this expression is the smallest possible.

28. Proof of Proposition 1 (cont-d)

- The expression $\max\limits_t |e_i(t)|$ is the L^∞-norm $\|e_i\|_{L^\infty}$.
- The expression $\int |e_i(s)|\,ds$ is the L_1-norm $\|e_i\|_{L^1}$.
- Thus, $J_{\rm int} = \Delta \cdot \|e_i\|_{L^\infty} \cdot \|e_i\|_{L^1}$.
- We consider the functions $e_i(t)$ for which $\int e_i^2(t)\,dt = 1$, i.e., $\|e_i\|_{L^2} = 1$, where $\|e_i(t)\|_{L^2} \stackrel{\rm def}{=} \sqrt{\int e_i^2(t)\,dt}$.
- There is a known Hölder's inequality connecting these three norms: $\|f\|^2_{L^2} \le \|f\|_{L^1} \cdot \|f\|_{L^\infty}$.
- 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

$$J_{\rm int} = \Delta \cdot \|e_i\|_{L^\infty} \cdot \|e_i\|_{L^1} \ge \Delta \cdot \|e_i\|^2_{L^2} = \Delta \cdot 1 = \Delta.$$

- It also implies that the smallest possible value Δ is attained when $|e_i(t)|$ is constant. Q.E.D.

29. Proof of Proposition 2

- It is known that $\int_a^b f(t)\,dt \le (b-a)\cdot \max\limits_s f(s)$.
- It is known that the equality happens only if $f(t) = \max\limits_s f(s)$ for almost all t.
- So, $\int_{\underline{T}_i}^{\overline{T}_i} e_i^2(t)\,dt \le (\overline{T}_i - \underline{T}_i)\cdot \max\limits_t e_i^2(t)$, and the equality is attained only if $|e_i(t)| = \text{const}$.
- For orthonormal $e_i(t)$, we have $\int_{\underline{T}_i}^{\overline{T}_i} e_i^2(t)\,dt = 1$.
- Thus, $\max\limits_t e_i^2(t) \ge \dfrac{1}{\overline{T}_i - \underline{T}_i}$, and the equality is attained if and only if $|e_i(t)| = \text{const}$.
- So, the minimum of $J_{\text{prob}}(e_i)$ is indeed attained when $|e_i(t)| = \text{const}$.
- The proposition is proven.