

Statistical analysis of a smoothing filter based on fuzzy transform

Michal Holčapek¹ Tomáš Tichý²

¹Institute for Research and Applications of Fuzzy Modeling,
University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic

²Department of Finance, Faculty of Economics VŠB-TU Ostrava,
Sokolská 33, 701 21 Ostrava 1, Czech Republic
michal.holcapek@osu.cz, tomas.tichy@vsb.cz

Abstract

The paper is devoted to the smoothing of discrete functions using the fuzzy transform (shortly, F-transform) introduced by Perfilieva in [1]. We generalize a smoothing filter based on the fuzzy transform proposed in [2] to obtain a better control on the smoothed functions. For this purpose, a generalization of the concept of fuzzy partition is suggested and the smoothing filter is defined as a combination of the direct discrete F-transform and a slightly modified inverse continuous F-transform. Statistical properties including the description of the white noise reduction and the asymptotic expression of Bias and Var are investigated and discussed.

Keywords: Fuzzy transform, smoothing filter, statistical analysis, white noise reduction

1. Introduction

In [2], we presented an application of F-transform to the non-parametric derivation of a probability density function (PDF) from a data sample. More precisely, we introduced an FT-smoothing filter which is the combination of the direct discrete and the inverse continuous F-transform and derived an optimization of parameters of uniform fuzzy partitions to obtain the best estimation of PDF with respect to the integrated square error (ISE). In comparison with the results obtained by the Parzen window estimator,¹ the FT-smoothing filter gives very similar estimates of PDFs, even for tails the fitting by the FT-smoothing filter seems to be better. In contrast to the Parzen window estimator, the FT-smoothing filter as well as the vector quantization based on Parzen windows, or the finite Gaussian mixture, can significantly decrease the model complexity being proportional for Parzen windows to the number of sample data and this can lead to the memory storage problem. A disadvantage of the FT-smoothing filter can be an overfitting for lower values of basic

functions bandwidth which is among others caused by a limitation of the original definition of the fuzzy partition.

In this paper, we generalize the FT-smoothing filter proposed in [2] to better check the smoothing of the resulted function. As we have mentioned above, the original definition of fuzzy partition has some limitations for the smoothing of discrete functions, since only two consecutive basic functions are important in the evaluation of the inverse F-transform. This fact has been pointed out by Stefanini in [5], where an extension of the fuzzy partition to a fuzzy r -partition² with $2r$ active basic functions³ and discrete and continuous $F^{(r)}$ -transform are suggested. Note that the sum of membership values over all basic functions for any point of the given interval is equal to r . Further, Stefanini uses advanced forms of fuzzy numbers for basic function modeling based on parametric shape functions (see [6]) and shows a smoothing effect on functions obtained from the $F^{(r)}$ -transform. Following the Stefanini's idea, we propose an extended fuzzy r -partition as a special of fuzzy r -cover, where we do not suppose that r is a natural number in general. Although, the fuzzy r -partitions are considered for any real number $r \geq 1$, practically, a request of a simple construction procedure based on Proposition 2.1 is to suppose r being a natural number. In this case, our definition is equivalent to Stefanini's definition and the values of r refer to the numbers of active basic function (see Remark 2.2).

In order to investigate statistical properties including asymptotic expression of Bias and Var in the fixed equally spaced design of the proposed FT-smoothing filter we use a close relation (see Proposition 3.1) between the components of F-transform and the Nadaraya-Watson (NW) estimator ([7, 8]) belonging among the traditional approaches to the kernel regression (see [3, 9, 10, 11]). Note that the asymptotic expressions are used for a comparison of the quality and rate of the convergence among filters based on kernels and, moreover, it is useful for the derivation of the asymptotic bandwidth h_{AMSE}

¹Note that the Parzen window estimator is one of the non-parametric methods as well as the histogram, vector quantization based on Parzen windows, or finite Gaussian mixture, using that one can estimate PDFs without any presumption on the shape and parameters of PDFs (for a survey, see [3, 4]).

²Note that r is a natural number greater than 0.

³That means a number of basic functions giving a non zero value in the evaluation of the direct F-transform.

to find an optimal model of an unknown function.

The contribution is structured as follows. The following section gives formal definitions of fuzzy r -cover, fuzzy r -partition and (uniform) fuzzy r -partition determined by a set of nodes and a set of so-called basal fuzzy sets. A one-to-one correspondence between basal fuzzy sets and kernels is proved. The third section is devoted to the definition of the direct discrete F-transform and to the comparison of the F-transform component as an estimator of unknown function values and the NW estimator. The main part is given in the fourth section, where a generalization of the FT-smoothing filter is proposed and its statistical properties are investigated and discussed. The last section is a conclusion.

2. Fuzzy r -partitions

Let \mathbb{R} denote the set of all real numbers. A *fuzzy set* on \mathbb{R} is a function $A : \mathbb{R} \rightarrow [0, 1]$. We say that a fuzzy set A is *empty*, if $A(x) = 0$ holds for any $x \in \mathbb{R}$. Further, we use $\ker(A)$ to denote the set of all $x \in \mathbb{R}$ for which $A(x) = 1$. The set $\ker(A)$ is usually called the *kernel* of A . We say that a fuzzy set is *convex*, if

$$A(\lambda x + (1 - \lambda)y) \geq \min(A(x), A(y))$$

for any $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$, *continuous*, if $A(x)$ is a continuous function in the common sense, and *normal*, if $\ker(A) = \{x\}$ for a suitable $x \in \mathbb{R}$. In this contribution, we will suppose that each fuzzy set is continuous, convex and has the non-empty kernel.⁴ Note that the choice of the fuzzy sets shapes is motivated by the shapes of fuzzy sets used in the original definition of the fuzzy partition in [1]. Finally, we will suppose that $R = [a, b]$.

Let us start with a more general concept than the fuzzy partition is called a fuzzy r -cover.

Definition 2.1 (Fuzzy r -cover). Let $r \geq 1$ be a real number. A *finite fuzzy r -cover* of R is a collection $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ of non-empty fuzzy sets $A_i : \mathbb{R} \rightarrow [0, 1]$ that satisfies

$$\sum_{i=1}^k A_i(x) \geq r \quad (1)$$

for any $x \in R$.

Note that the condition (1) is a consequence of the natural generalization of a finite cover of a set by sets, when the disjointness of different sets is not supposed in general. In the following text, we will omit “finite” in the “finite fuzzy r -cover”, since the other than a finite fuzzy r -cover is not considered here. A simple but very useful consequence of the definition of the fuzzy r -cover is as follows.

⁴Note that if $\text{Supp}(A) = \text{cl}\{x \in \mathbb{R} \mid A(x) > 0\}$, where cl is the closure operator, is bounded, one can see a special type of fuzzy intervals or especially fuzzy numbers (cf. [12]).

Proposition 2.1. Let \mathcal{A}_1 and \mathcal{A}_2 be disjoint fuzzy r_1 -cover and r_2 -cover of R , i.e. $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$. Then $\mathcal{A}_1 \cup \mathcal{A}_2$ is a fuzzy $(r_1 + r_2)$ -cover of R .

Using the concept of the fuzzy r -cover we can naturally generalize the concept of the fuzzy partition of R . We use $A \upharpoonright R$ to denote the restriction of a function A to R .

Definition 2.2 (Fuzzy r -partition). A *finite fuzzy r -partition* of R is a collection $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ of non-empty fuzzy sets such that there exists a fuzzy r -cover $\mathcal{B} = \{B_i \mid i = 1, \dots, k\}$ of R such that

1. $A_i = B_i \upharpoonright R$ for any $i = 1, \dots, k$,
2. $\sum_{i=1}^k A_i(x) = r$ for any $x \in R$.

The fuzzy sets of \mathcal{A} are called the *basic functions*.

Simply speaking, a fuzzy r -partition of R is the restriction of functions of a special fuzzy r -cover to R , where a generalization of the “Ruspini condition” is considered.⁵ In a comparison with the original definition in [1], the basic functions in our conception are defined over nodes that do not belong to R in general. This enables us to better regulate the smoothness of resulted functions (cf. [5]).

Remark 2.1. Obviously, one can reformulate Proposition 2.1 for the fuzzy r -partitions to construct more complex fuzzy r -partitions from the simple ones. Note that this procedure has been used in the construction of fuzzy r -partitions in [6], where each fuzzy r -partition can be derived as the combination of fuzzy 1-partitions.

In the following part, we are mainly interested in a special case of fuzzy r -partitions, namely, a *uniform fuzzy r -partition determined by a basal fuzzy set*. A main reason for this step is the study of the asymptotic properties of the FT-smoothing filter estimator in the same spirit as is done for the kernel based smoothing estimators.

Definition 2.3. A *basal fuzzy set* is a continuous, convex, normal fuzzy set $S : \mathbb{R} \rightarrow [0, \infty)$ such that $S(x) = S(-x)$ for any $x \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty} x^2 S(x) dx < \infty.$$

One can see that a basal fuzzy set is a symmetric fuzzy set satisfying $\int_{-\infty}^{\infty} x S(x) dx = 0$. Note that this equality enables us to significantly simplify the description of e.g. the Bias, Var, or AMSE being intensively investigated in the non-parametric regression theory (see [9]).

Definition 2.4. Let R be a real interval, $\mathcal{T} = \{t_i \mid i = 1, \dots, k\}$ be an ordered set of nodes from \mathbb{R} with $t_i < t_{i+1}$ for any $i = 1, \dots, k - 1$ and $\mathcal{S} =$

⁵Recall that a fuzzy partition $\{A_1, \dots, A_n\}$ of R satisfies the Ruspini condition, if $\sum_{i=1}^n A_i(x) = 1$ for any $x \in R$.

$\{S^{(i)} \mid i = 1, \dots, k\}$ be a set of basal fuzzy sets. A fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$ is a fuzzy r -partition $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ of R such that

$$A_i(x) = S^{(i)}(x - t_i) \quad (2)$$

holds for any $x \in R$ and $i = 1, \dots, k$. We say that a fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$ is *uniform*, if $S = S^{(i)}$ for any $i = 1, \dots, k$ and $t_{i+1} - t_i = u$ for any consecutive nodes t_{i+1} and t_i in \mathcal{T} .

In the case, when a fuzzy r -partition determined by $(\mathcal{T}, \mathcal{S})$ is uniform and $S^{(i)} = S$ for all $i = 1, \dots, k$, we will also write (\mathcal{T}, S) instead of $(\mathcal{T}, \mathcal{S})$.

Remark 2.2. An interesting property showing something about the uniform fuzzy r -partitions says that if r is a natural number, $u = t_{i+1} - t_i$ and h denotes the bandwidth of a basal fuzzy set, then $ru = h$. Note that an open problem is to find a uniform fuzzy r -partition for a real value of r which is not natural number and eventually to prove an analogous relation among parameters r , h and u .

The following examples of uniform fuzzy r -partitions are based on the basal fuzzy sets (cf. [1]). For interesting examples of non-uniform fuzzy r -partitions based on parameterized fuzzy numbers, we refer to [5].

Example 2.3 (Triangle fuzzy r -partition). Let us consider a triangle basal fuzzy sets given by

$$S_h(x) = \max\left(0, \frac{h - |x|}{h}\right) \quad (3)$$

and $R = [0, 10]$. Using Proposition 2.1, one can simply verify that fuzzy r -partitions presented on Fig.1 are a uniform fuzzy 2-partition of R with the bandwidth $h = 2$ and a uniform fuzzy 9-partition of R with the bandwidth $h = 6$. For a better orientation, we split the uniform fuzzy r -partitions onto two parts pictured by normal and dashed lines that are again fuzzy r -partitions.

Example 2.4 (Raised Cosine fuzzy r -partition). Let us consider a raised cosine basal fuzzy set given by

$$S_h(x) = \begin{cases} \frac{1}{2} \left(1 + \cos\left(\frac{x}{h}\pi\right)\right), & x \in [-h, h]; \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

and $R = [0, 10]$. On Fig. 2, we can see a uniform fuzzy 2-partition and a uniform fuzzy 4-partition of R with the same bandwidth $h = 4$. For a verification, we again split these fuzzy r -partitions onto two parts pictured by normal and dashed lines.

In order to introduce the NW estimator playing an important role in our study of statistical properties of FT-smoothing filter, let us establish the

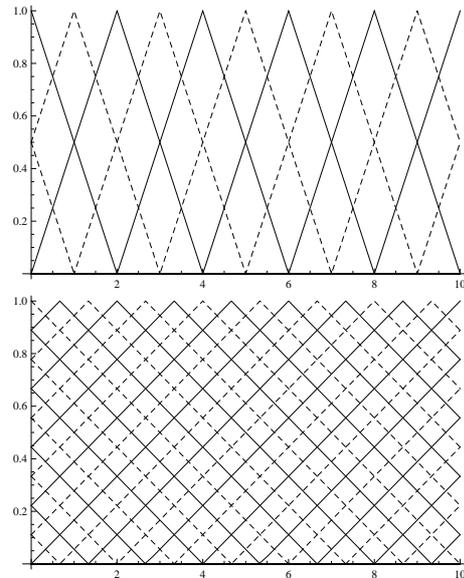


Figure 1: Triangle uniform fuzzy 2-partition (left) for $h = 2$ and 9-partition (right) for $h = 6$.

concept of kernel. A *kernel* is a continuous, symmetric, unimodal function $K : \mathbb{R} \rightarrow [0, \infty)$ having the following properties:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \text{ and } \int_{-\infty}^{\infty} x^2 K(x)dx < \infty.$$

Obviously, each kernel is a symmetric density function. Typical examples of kernels are the Uniform, Triangular, Raised cosine, Epanechnikov or Gaussian density functions (see [13, 3, 11]). Let $h > 0$ be a real number and K be a kernel, then $K_h(x) = 1/hK(x/h)$ is again a kernel. The value of h is also called the *bandwidth* (see [9]). For our investigation, the following relation between basal fuzzy sets and kernels is fundamental.

Proposition 2.2. *There is a one-to-one correspondence between the classes of all basal fuzzy sets and all kernels.*

Proof. Defining $F(S)(x) = S(x) / \int_{-\infty}^{+\infty} S(x)dx$ and $G(K)(x) = K(x)/K(0)$, then one can simply check that $F \circ G(K) = K$ and $G \circ F(S) = S$. \square

3. Direct discrete fuzzy transform

Let $\text{Dom}(g)$ denote the domain of a function g , $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ be a fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$ and $s \in \{1, 2, \dots\}$. We say that a set $X = \{x_j \mid j = 1, \dots, n\}$ of reals is *s-dense with respect to \mathcal{A}* , if for each $A_i \in \mathcal{A}$ there exist at least s nodes $x_{j_1}, \dots, x_{j_s} \in X$ such that $A_i(x_{j_t}) > 0$ holds for any $t = 1, \dots, s$. If the set X is 1-dense with respect to \mathcal{A} , we say that X is *sufficiently dense with respect to \mathcal{A}* .⁶ Note that the concept of “being sufficiently dense” has been introduced in [1].

⁶Usually, it is sufficient to suppose that sets of nodes are 1-dense. Nevertheless, there are some cases when the pre-

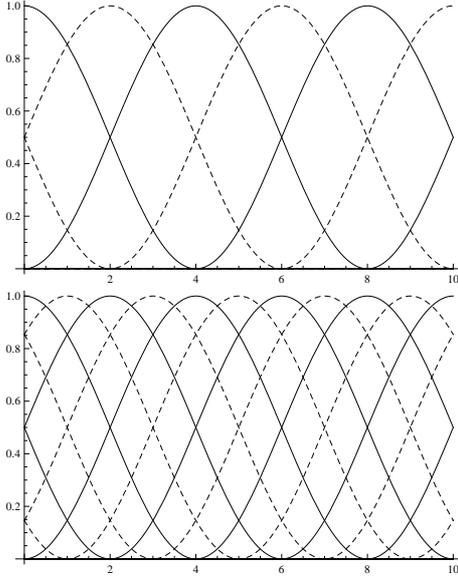


Figure 2: Raised cosine uniform fuzzy 2-partition (left) and 4-partition (right) for $h = 4$.

Definition 3.1 ([1]). Let g be a finite real function given at the nodes $x_1 < \dots < x_n$ with $\text{Dom}(g) \subseteq R$ and $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ be a fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{K})$ such that $\text{Dom}(g)$ is sufficiently dense with respect to \mathcal{A} . We say that a collection of real numbers $\{F_i \mid i = 1, \dots, k\}$ is the *discrete fuzzy (F-)transform of g with respect to \mathcal{A}* , if

$$F_i = \frac{\sum_{j=1}^n g(x_j) A_i(x_j)}{\sum_{j=1}^n A_i(x_j)}. \quad (5)$$

The numbers F_i are called *components of the discrete F-transform*.

Let us make useful simplifications and presumptions. First, we will omit “discrete” in the “discrete F-transform”. Further, we will omit “determined by $(\mathcal{T}, \mathcal{S})$ ” in the “fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$ ”, when it does not induce any confusion. Finally, we will suppose that each given set of nodes, at which a function is given, is sufficiently dense with respect to a given fuzzy r -partition.

Notice that the presumption of “being sufficiently dense” used in the previous definition ensures the correctness of the formula (5), i.e. the denominator is different from 0 for each basic function A_i . Since A_i is determined by a basal fuzzy set $S^{(i)}$, then we may also write

$$F_i = \frac{\sum_{j=1}^n g(x_j) S^{(i)}(x_j - t_i)}{\sum_{j=1}^n S^{(i)}(x_j - t_i)}, \quad (6)$$

or even

$$F_i = \frac{\sum_{j=1}^n g(x_j) S(x_j - t_i)}{\sum_{j=1}^n S(x_j - t_i)}, \quad (7)$$

sumption of s -density for $s > 1$ leads to a better result (see Corollary 5.3).

if a uniform fuzzy r -partition is supposed.

Let us present some statistical properties of components of F-transform. Before commencing our study, let us give some relevant terminology and notation. A *fixed design* consists of x_1, \dots, x_n which are ordered non-random numbers. We shall say that a fixed design is *equally spaced*, if $x_{j+1} - x_j$ is constant for all $j = 1, \dots, n$. Sometimes, a fixed equally spaced design has a form $x_j = \frac{j}{n}$ for $j = 0, \dots, n$. For the fixed design case the response variable are assumed to satisfy

$$Y_j = m(x_j) + \varepsilon_j, \quad j = 1, \dots, n \quad (8)$$

where m is a (non-random) function and ε_j is a random variable representing the error in x_j having the following properties $E(\varepsilon_j) = 0$, $\text{Var}(\varepsilon_j) < \infty$ and $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. In our investigation, we restrict ourselves to the common choice $\text{Var}(\varepsilon_j) = \sigma^2$ for any $j = 1, \dots, n$ (a homoscedastic model). One can prove easily that $E(Y_j) = m(x_j)$, $\text{Var}(Y_j) = \text{Var}(\varepsilon_j) = \sigma^2$ and $\text{Cov}(Y_i, Y_j) = 0$ for any $i \neq j$.⁷

The aim of the kernel-based nonparametric regression is to estimate the unknown function m . There are many methods based on one or more than one kernels how to find a “good” estimation of m (see [13, 14, 3, 10, 4, 11, 9]). Here, we restrict ourselves to one of them called the Nadaraya-Watson (NW) estimator introduced in [7, 8]:

$$\hat{m}(x) = \frac{\sum_{j=1}^n Y_j K(x_j - x)}{\sum_{j=1}^n K(x_j - x)}, \quad (9)$$

where Y_j is expressed by (8). Comparing the NW estimator with the formula (5) for the F-transform components, where we consider random variables Y_j instead of values $g(x_j)$, one can see the similarity. More precisely, we may state the following proposition.

Proposition 3.1. *Let R be a real interval and $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ be a uniform fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$. If y_1, \dots, y_n are values of random variables Y_1, \dots, Y_n defined by (8) and \hat{m} is the NW estimator with $K(x) = S(x) / \int_{-\infty}^{+\infty} S(x) dx$, then $F_i = \hat{m}(t_i)$ for any $t_i \in \mathcal{T}$.*

Let $R = [a, b]$ be an interval and consider the fixed equally spaced design regression model

$$Y_j = m(x_j) + \varepsilon_j, \quad j = 1, \dots, n, \quad (10)$$

where $x_j = a + j \left(\frac{b-a}{n}\right)$ and $\text{Var}(\varepsilon_j) = \sigma^2$. Note that one could also consider $x_0 = a$, but $x_0 \notin d_{\mathcal{A}}(x)$ for any $x \in [a+h, b-h]$ (see (A4) below) and thus x_0 can be omitted in our analysis. We shall make the following presumptions in our analysis (cf. [9]):

(A1) The function m'' is continuous on $[a, b]$.

⁷For example, $\text{Cov}(Y_i, Y_j) = E((Y_i - m(x_i))(Y_j - m(x_j))) = E((\varepsilon_i - 0)(\varepsilon_j - 0)) = \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$.

- (A2) K is a symmetric kernel with $K(x) = 0$ for any $x \notin (-1, 1)$.
- (A3) The bandwidth $h = h_n$ is a sequence satisfying $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$.
- (A4) The point x at which the estimation is taking place satisfies $a + h_n < x < b - h_n$ for all $n_0 \leq n$ where n_0 is fixed.

Note that usually $[a, b]$ in the fixed equally spaced design context is considered to be the unit interval and recall that $\text{Bias}(\hat{m}(x)) = E(\hat{m}(x)) - m(x)$.

Theorem 3.2. *Under the presumptions (A1)-(A4), we may write*

$$\begin{aligned} \text{Bias}(\hat{m}(x)) &= \frac{m''(x)h^2\mu_2(K)}{2} + o(h^2) + O\left(\frac{1}{n}\right), \\ \text{Var}(\hat{m}(x)) &= \frac{\sigma^2(b-a)}{nh}R(K) + o\left(\frac{1}{nh}\right), \end{aligned}$$

where we put $\mu_2(K) = \int_{-1}^1 z^2 K(z) dz$ and $R(K) = \int_{-1}^1 K^2(z) dz$.

A proof of this theorem can be found in e.g. [9, 13]. One can see that $\hat{m}(x)$ is an asymptotically unbiased estimator of m . A simple consequence of this theorem is a computation of *mean square error*

$$\begin{aligned} \text{MSE}(\hat{m}(x)) &= \text{Var}(\hat{m}(x)) + \text{Bias}(\hat{m}(x))^2 = \\ &= \frac{\sigma^2(b-a)}{nh}R(K) + \frac{(m''(x))^2 h^4 \mu_2(K)^2}{4} + \\ &= o\left(\frac{1}{nh} + h^4\right). \end{aligned}$$

Let us denote

$$\begin{aligned} \text{AMSE}(\hat{m}(x)) &= \\ &= \frac{\sigma^2(b-a)}{nh}R(K) + \frac{(m''(x))^2 h^4 \mu_2(K)^2}{4} \end{aligned}$$

the asymptotic MSE. The optimal value of bandwidth h can be derived putting to zero the derivative of AMSE with respect to h (for details, we refer to [9, 13]). By a simple calculation we obtain

$$h_{\text{AMSE}}^{\text{NW}} = \left(\frac{\sigma^2(b-a)}{nm''(x)} C(K) \right)^{\frac{1}{5}} \quad (11)$$

where $C(K) = \frac{R(K)}{\mu_2(K)^2}$ may be understood as a characterization of the kernel K .

4. FT-smoothing filter

Let \mathcal{A} be a fuzzy r -partition of R . We will use $F(R, \mathcal{A})$ to denote the set of all functions g such that $\text{Dom}(g) \subseteq R$ and $\text{Dom}(g)$ is sufficiently dense with respect to \mathcal{A} . Obviously, the set $F(R, \mathcal{A})$ contains all function on which the discrete F-transform may be applied. Further, we will $\text{CF}(R, \mathcal{A})$ to denote the set of all continuous real functions g defined on R .

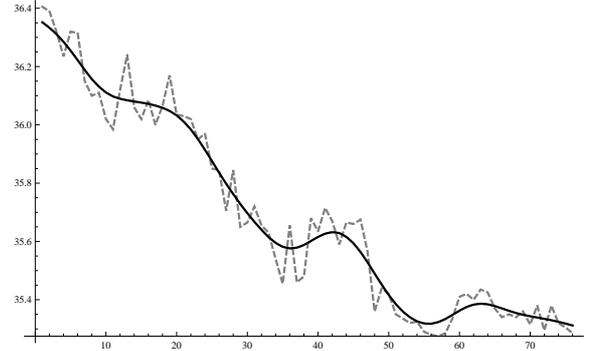


Figure 3: Application of FT-smoothing filter on a geometric Brownian motion ($h = 5$ and $r = 5$).

Definition 4.1 (FT-smoothing filter). Let R be a real interval and $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ be a fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$. An *FT-smoothing filter determined by \mathcal{A}* is a mapping $\mathcal{F}_{\mathcal{A}} : F(R, \mathcal{A}) \rightarrow \text{CF}(R, \mathcal{A})$ defined by

$$\mathcal{F}_{\mathcal{A}}(g)(x) = \frac{1}{r} \sum_{i=1}^r F_i A_i(x) \quad (12)$$

for any $x \in R$, where F_k are the components of the discrete F-transform.

Note that the formula (12) has been used in [5] for the inverse discrete and continuous $F^{(r)}$ -transform. On Fig. 3, one can see an application of the FT-smoothing filter (based on a triangle basal fuzzy set) to the smoothing of a geometric Brownian motion in finance to obtain a better imagination about the structure of this process. Since the results for the triangle and raised cosine types of basal fuzzy sets are nearly identical, we restrict ourselves in this paper to the triangle type of basal fuzzy sets and the fuzzy r -partitions determined by them.

5. Statistical analysis of FT-smoothing filter

In this part, we will apply the results of our statistical analysis of the F-transform components to the statistical analysis of FT-smoothing filter.

Let us suppose a fixed design, i.e., $x_1 < \dots < x_n$ is a sequence of (non-random) nodes of R . Further, let \mathcal{A} be a fuzzy r -partition of R determined by $(\mathcal{T}, \mathcal{S})$. Further, let us suppose that $X = \{x_1, \dots, x_n\}$ is sufficiently dense with respect to \mathcal{A} . Finally, put $g(x_j) = Y_j = m(x_j) + \varepsilon_j$, $j = 1, \dots, n$, and denote

$$\hat{m}_{FT}(x) = \mathcal{F}_{\mathcal{A}}(g)(x) \quad (13)$$

the estimator of function $m(x)$ based on the FT-smoothing filter. The following theorem expresses the Bias and Var of the estimated function $\hat{m}_{FT}(x)$.

Theorem 5.1. Let $x \in R$ and $\hat{m}(x)$ denote the NW estimator of m at the point x . Then we may write

$$\begin{aligned} \text{Bias}(\hat{m}_{FT}(x)) &= \frac{1}{r} \sum_{i=1}^k \text{Bias}(\hat{m}(t_i)) A_i(x) + \\ &\quad \frac{1}{r} \sum_{i=1}^k (m(t_i) - m(x)) A_i(x), \\ \text{Var}(\hat{m}_{FT}(x)) &= \frac{1}{r^2} \sum_{i=1}^k \sum_{j=1}^k \text{Cov}(F_i, F_j) A_i(x) A_j(x). \end{aligned}$$

Note that the expressions of Bias and Var have the same form also for the random design. The Bias shows that the estimator based on the FT-smoothing filter is biased in both designs analogously to the kernel based estimators. For the fixed design, an expression of $\text{Cov}(F_i, F_j)$ in terms of basic functions can be done as follows. Unfortunately, there are some problems in the equivalent expression of Bias for the random design and this expression is an open problem.

Theorem 5.2. Let $x \in R$. Then we may write

$$\text{Var}(\hat{m}_{FT}(x)) = \frac{\sigma^2}{r^2} \sum_{i=1}^k \sum_{j=1}^k a_{ij} A_i(x) A_j(x),$$

where

$$a_{ij} = \frac{\sum_{s=1}^n A_i(x_s) A_j(x_s)}{\sum_{s=1}^n A_i(x_s) \sum_{s=1}^n A_j(x_s)}.$$

A simple but very interesting consequence of the previous description of $\text{Var}(\hat{m}_{FT}(x))$ is the following inequality saying that under some conditions the FT-smoothing filter really reduces the white noise.

Corollary 5.3. If $X = \{x_1, \dots, x_n\}$ be a set of nodes that is 2-dense with respect to A . Then $\text{Var}(\hat{m}_{FT}(x)) < \sigma^2$.

In the rest of this section, we will study the asymptotic properties of the estimator based on the FT-smoothing filter in the fixed equally spaced design. In order to use the results about the asymptotic behavior of the NW estimator, we will express the FT-smoothing filter in terms of kernels (see Proposition 2.2). One can simply verify that if we consider a uniform fuzzy r -partition determined by (\mathcal{T}, S) for $S(x) = K(x)/K(0)$, where K is the kernel determined by S , then

$$\mathcal{F}_A(g)(x) = \frac{1}{rK(0)} \sum_{i=1}^k F_i K(x - t_i).$$

Since $\sum_{j=1}^n K(x - t_j) = rK(0)$ for any $x \in R$, we may introduce equivalently a uniform kernel r' -partition determined by (\mathcal{T}, K) with $r' = rK(0)$ and adopt all properties defined for the uniform fuzzy r -partitions (e.g. the s -density of a set with respect to a uniform kernel r -partition). Further,

we may introduce the FT-smoothing filter in terms of kernels by

$$\mathcal{F}_A(g)(x) = \frac{1}{r'} \sum_{i=1}^k F_i K(x - t_i),$$

where the components F_i are equal to $\hat{m}(t_i)$ derived by the NW estimator (see Proposition 3.1). Let us stress that the interpretation of values of r' is different from the interpretation of values of r for the fuzzy r -partitions. While the value of r refers more or less to a number of active basic functions (especially, expressed by basal fuzzy sets) for an element $x \in R$ over which the evaluation of FT-smoothing filter is made (cf. [5]), the value of r' is only an abstract value strongly dependent on the high of the value $K(0)$. For simplicity, we will omit the comma in r' and write only r in the following text.

Before commencing our study of asymptotic properties of the FT-smoothing filter (equivalently expressed by kernels) in the fixed equally spaces design, let us recall that $Y_j = m(x_j) + \varepsilon_j$, $j = 1, \dots, n$, where $x_j = a + j(b - a)/n$, m is a (non-random) function and ε_j is a random variable representing the error in x_j with $E(\varepsilon_j) = 0$, $\text{Var}(\varepsilon_j) = \sigma^2 < \infty$, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. Further, let us denote \mathcal{A}_k a uniform kernel r_k -partition of R determined by (\mathcal{T}_k, K_h) , where k is the number of nodes over which the uniform kernel r_k -partition is constructed. Finally, let us define

$$\alpha(x, y) = \int_{-1}^1 K(z) K\left(\frac{x - y}{h} + z\right) dz$$

and put $u_k = u = t_{i+1} - t_i$. Obviously, $\alpha(x, x) = R(K)$. Notice that $u \leq h$, otherwise, \mathcal{A}_k cannot be a kernel r -partition since $\sum_{i=1}^k A_i(x) = 0$ for some $x \in [a, b]$.

To show the asymptotic behavior of FT-smoothing filter, we will make the following assumption in our analysis:

- (FA1) The functions m is continuously differentiable up to the fourth order on $[a, b]$.
- (FA2) K is a symmetric kernel with $K(x) = 0$ for any $x \notin (-1, 1)$.
- (FA3) The bandwidth $h = h_n$ is a sequence satisfying $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$.
- (FA4) The point x at which the estimation is taking place satisfies $a + 2h_n < x < b - 2h_n$ for all $n \geq n_0$, where n_0 is fixed.
- (FA5) \mathcal{A}_k is a kernel r -partition of $[a + h_n, b - h_n]$ for all $k \geq k_0$, where k_0 is fixed.
- (FA6) $r = r_k$ and $u = u_k$ are sequences satisfying $r_k \rightarrow \infty$, $nh_n^3 k^2 \rightarrow \infty$, $\frac{h_n^2}{r_k u} \rightarrow 0$ and $\frac{k^2}{r_k^2 n} \rightarrow 0$ as $n \rightarrow \infty$ and $k \rightarrow \infty$.
- (FA7) $\alpha(x, y)$ has continuous partial derivatives in $[a, b]$.

Let us only say that the mentioned convergences are important presumptions under which the asymptotic expressions of Bias and Var of $\hat{m}_{FT}(x)$ converge to zero. An asymptotic characterization of Bias and Var of the estimator $\hat{m}_{FT}(x)$ can be described as follows.

Theorem 5.4. *Let the presumptions (FA1)-(FA7) be satisfied and $\frac{1}{n} \in O\left(\frac{1}{r}\right)$, then*

$$\begin{aligned} \text{Bias}(\hat{m}_{FT}(x)) &= \\ & \frac{h^2 \mu_2(K)}{ru} m''(x) + o\left(\frac{h^2}{ru} + h^2\right) + O\left(\frac{1}{r}\right), \\ \text{Var}(\hat{m}_{FT}(x)) &= \\ & \frac{k^2 \sigma^2}{r^2 n h (b-a)} R(K) + o\left(\frac{k^2}{nr^2} + \frac{1}{nh^3 r^2}\right), \end{aligned}$$

where $\mu_2(K)$, $R(K)$ are defined in Theorem 3.2.

The presumptions on $k \rightarrow \infty$, the bandwidth $h \rightarrow 0$ and $h^2 r^{-1} u^{-1} \rightarrow 0$ implies that $\hat{m}_{FT}(x)$ is an asymptotically unbiased estimator of m similarly to the NW estimator. Let us suppose that $k = k_n \rightarrow \infty$ and $h = h_n \rightarrow 0$ are sequences dependent on the number of data n such that $r_{k_n} u_{k_n} / K_{h_n}(0) = h_n$ (clearly, the bandwidth h_n is the same for a basal fuzzy set and its kernel counterpart). Note that, according to Remark 2.2, this equality is automatically satisfied for the values of r_{k_n} being natural numbers. Since $K_h(0) = K(0)/h$, then omitting the indexes for simplicity, one may derive $ru/K(0) = 1$ and thus $ru = K(0)$. Hence, we obtain another form of Bias of $\hat{m}_{FT}(x)$

$$\begin{aligned} \text{Bias}(\hat{m}_{FT}(x)) &= \frac{h^2 \mu_2(K)}{K(0)} m''(x) + \\ & o\left(\frac{h^2}{K(0)} + h^2\right) + O\left(\frac{1}{r}\right). \end{aligned}$$

In comparison with the Bias of $\hat{m}(x)$ obtained by the NW estimator, one can see that the rate of the convergence in the case of the FT-smoothing filter is slower than in the case of the NW estimator, since $\frac{h^2}{K(0)} + h^2 \notin o(h^2)$ and, clearly, $1/r \notin O(1/n)$, otherwise, the original idea of partitioning of intervals is missing. On the other hand, the model complexity for the FT-smoothing filter may be significantly smaller than for the NW, if a higher number of data is considered.

Since the MSE has a rather complicated form (mainly due to the expression of limiting behavior described by the big O notation and the little o notation), we omit it here. However, the AMSE has the form

$$\begin{aligned} \text{AMSE}(\hat{m}_{FT}(x)) &= \\ & \frac{k^2 \sigma^2}{r^2 n h (b-a)} R(K) + \frac{h^4 \mu_2(K)^2}{r^2 u^2} m''(x)^2. \end{aligned}$$

The optimal value of bandwidth h_{AMSE} can be derived putting to zero the derivative of AMSE with

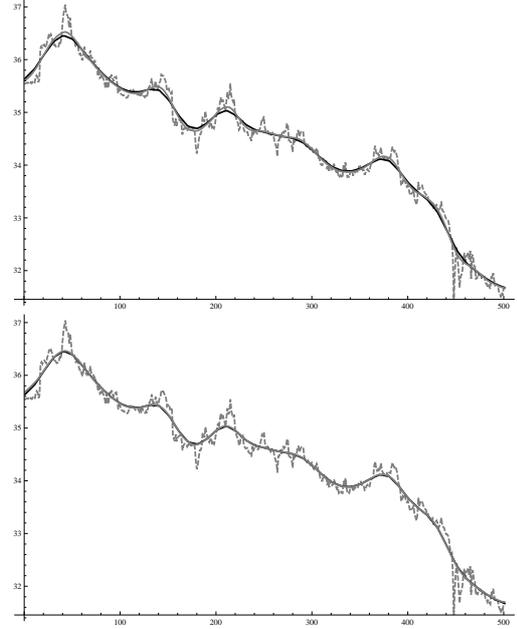


Figure 4: Comparison of $h^{\text{NW}} = h^{\text{FT}}$ (left) and $h^{\text{NW}} = 1/0.76h^{\text{FT}}$ (right) for $h^{\text{FT}} = 20$ and $r = 2$.

respect to h . By a simple calculation we obtain

$$h_{\text{AMSE}}^{\text{FT}} = \left(\frac{k^2 u^2 \sigma^2}{4n m''(x)^2 (b-a)} C(K) \right)^{\frac{1}{5}}$$

where $C(K) = R(K)/\mu_2(K)^2$. Recall that $u = t_{i+1} - t_i$ and k denotes the number of basic functions obtained by the kernel K . Since $t_i \notin R$ for some $i = 1, \dots, k$, we can simply deduce $ku \geq b-a$. One can notice that $ku \approx b-a$ for small h and

$$h_{\text{AMSE}}^{\text{FT}} \approx \left(\frac{(b-a)\sigma^2}{4n m''(x)^2} C(K) \right)^{\frac{1}{5}}.$$

Comparing this result with that provided by the NW estimator, we obtain an approximated equality

$$h_{\text{AMSE}}^{\text{FT}} \approx 0.76 h_{\text{AMSE}}^{\text{NW}}, \quad (14)$$

where $0.76 \approx \sqrt[5]{1/4}$. Thus, a lower asymptotic bandwidth for the FT-smoothing filter is needed to obtain an optimal model of an unknown function which would correspond to the NW model.

On Fig. 4, one can see a comparison⁸ of the resulted functions obtained by the NW estimator (gray line) and by the FT-smoothing filter (black line), where we use $h^{\text{NW}} = h^{\text{FT}}$ (top) and $1/0.76h^{\text{NW}} = h^{\text{FT}}$ (bottom) with $h^{\text{FT}} = 20$ and $r = 2$. Although, a small number of h is not here supposed, the correction of h^{FT} against h^{NW} works well. On Fig. 5, we demonstrate the smoothing property which is clearly dependent on the level of r for fixed values $h^{\text{FT}} = 20$ and $h^{\text{NW}} = 20/0.76$. The

⁸For this purpose, we chose data set consisting of 500 daily quotas of CZK/EUR exchange rate randomly selected from the original time series covering the last eight years.

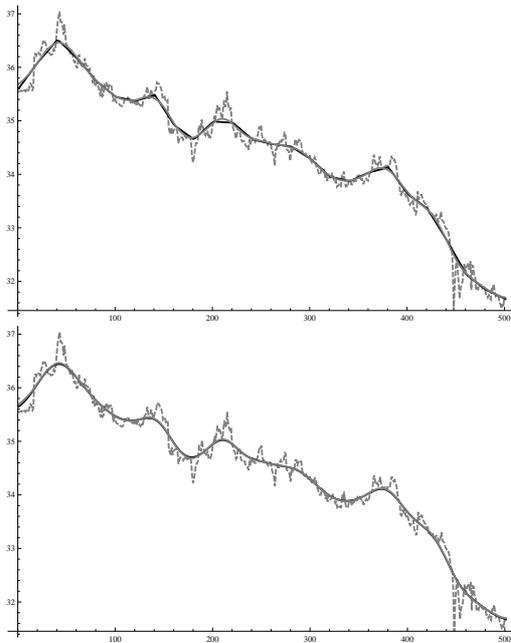


Figure 5: Comparison of the smoothness of resulted functions for $h^{\text{NW}} = 1/0.76h^{\text{FT}}$ with $h^{\text{FT}} = 20$ and $r = 1$ (left) and $r = 8$ (right)

both approaches give practically identical resulted functions. Note that 26 components are used and only 2 basic functions are active (giving a non-zero value) for $r = 1$ to find the values of the smoothed function. For $r = 8$, the number of components and basic functions used for calculation is naturally much greater, namely, 215 components and at most 16 basic functions. On the other hand, this is still less than the number of values over which is the kernel active, namely, 40 values are used to evaluate the function values using NW estimator. Summarizing our observation, the FT-smoothing filter can be advantageously used in cases when higher numbers of data are considered to reduce the model complexity of NW estimator, but to retain the quality of estimates.

6. Conclusion

The paper was devoted to a smoothing technique that generalizes the one proposed in [2] and it is based on the F-transform. To obtain better results of the smoothing procedure, analogously as in [5], we extended the concept of fuzzy partition to the fuzzy r -partition, where the sum of function values for basic functions may be greater than or equal to 1, i.e., the Ruspini condition is not satisfied here in general. Over such fuzzy r -partitions, the components of the direct discrete F-transform and the FT-smoothing filter were introduced. The FT-smoothing filter is defined as a combination of the direct discrete F-transform introduced in [1] and a formula for the inverse continuous F-transform proposed in [5]. An interesting relation between the

components of the F-transform and the values of the NW estimator was proved and used in the study of statistical properties of this filter as the reduction of the white noise, Bias, Var and AMSE. The theoretical results were commented in some details and a comparison of the FT-smoothing filter and NW estimator was demonstrated by figures.

Acknowledgement

The work on this paper was partly supported by the Institutional Research Plan MSM 619 8898 701 and the SGS project SP2011/7 of VŠB-TU Ostrava.

References

- [1] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets syst.*, 157(8):993–1023, 2006.
- [2] M. Holčápek and T. Tichý. A probability density function estimation using F-transform. *Kybernetika*, 46(3):447–458, 2010.
- [3] W. Hölder, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and semiparametric models*. Springer-Verlag, Berlin Heidelberg, 2004.
- [4] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, London, 1986.
- [5] L. Stefanini. Fuzzy transform and smooth function. In *Proceedings of IFSA/EUSFLAT 2009*, pages 579–584, Lisbon, Portugal, 2009.
- [6] L. Stefanini, L. Sorini, and M.L. Guerra. Parametric representation of fuzzy numbers and application to fuzzy calculus. *Fuzzy Sets Syst.*, 157(18):2423–2455, 2006.
- [7] E.A. Nadaraya. On estimating regression. *Theory Probab. Appl.*, 9(1):141–142, 1964.
- [8] G.S. Watson. Smooth regression analysis. *Sankhya*, Series A, 26:359–372, 1964.
- [9] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman&Hall/CRC Monographs on Statistics & Applied Probability, London, 1995.
- [10] A. Pagan and A. Ullah. *Nonparametric Econometrics*. Cambridge University Press, New York, 1999.
- [11] J. S. Simonoff. *Smoothing methods in statistics*. Springer-Verlag, New York, 1996.
- [12] D. Dubois and H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9:613–626, 1978.
- [13] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman&Hall/CRC Monographs on Statistics & Applied Probability, 1996.
- [14] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics. Springer-Verlag, 2005.