Fuzzy sets and systems. 2011, vol. 180, issue 1, p. 69-97. http://dx.doi.org/10.1016/j.fss.2011.05.028

A Smoothing Filter Based on Fuzzy Transform☆

Michal Holčapek[a,*], Tomáš Tichý[b]

[a]*Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic*
[b]*Department of Finance, Faculty of Economics VŠB-TU Ostrava, Sokolská 33, 701 21 Ostrava 1, Czech Republic*

Abstract

The paper is devoted to the smoothing of discrete functions using the fuzzy transform introduced by Perfilieva. We generalize a smoothing filter based on the fuzzy transform recently proposed by us 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 fuzzy transform and a slightly modified inverse continuous fuzzy transform. An approximation behavior, total variation of smoothed functions and statistical properties including the description of the white noise reduction and the asymptotic expression of bias and variance are investigated and discussed. The proposed filter is compared with the Nadaraya-Watson estimator and the results are illustrated assuming financial data.

Key words: fuzzy transform, fuzzy partition, Nadaraya-Watson estimator, kernel regression, noise reduction, financial returns

1. Introduction

In statistics and image or signal processing, to smooth a data set means to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena. In literature, we can find many different smoothing and filter types that are based, for example, on the stochastic processes, kernel regressions, integral or wavelet transforms, or on the techniques of the fuzzy set theory (see e.g. [10, 15, 16, 17, 19, 28, 34]). Obviously, nobody can imagine today's life without some smoothing techniques, for example, a noise reduction in television or radio signals to improve the quality of telecast or broadcast, respectively, sometimes

☆The paper has been partially supported by the Institutional Research Plan MSM 6198898701 and the project of GAČR under No. 402/08/1237.

*Corresponding author

Email addresses: michal.holcapek@osu.cz (Michal Holčapek), tomas.tichy@vsb.cz (Tomáš Tichý)

a defect reduction in photographs can be used to improve the quality of pictures, or an outlier reduction in financial time series to better understanding of financial process behavior .

The fuzzy transform (shortly, F-transform) is a simple approximation technique proposed by Perfilieva in [24] (see also [22]) based on fuzzy partitioning of a closed real interval into fuzzy subsets. The F-transform is introduced and investigated in a continuous and discrete design context. In both designs, a (continuous or finite) function defined on a closed interval $[a, b]$ is transformed, using fuzzy sets (basic functions) that form a fuzzy partition of $[a, b]$, to a finite number of real numbers called the components of F-transform. This type of F-transform is called the direct F-transform. An inverse F-transform assigns to F-transform components a continuous function in the continuous design and a finite function in the discrete design, which is an approximation of the original function. A simplification and a good approximation of the original function makes from the F-transform a powerful tool that can be applied, for example, in data analysis or image processing (see e.g. [2, 4, 5, 6, 23, 32, 26, 33]). A generalization of F-transform based on the strict continuous triangular norms can be found in [3]. An interesting extension of the F-transform technique to the case where its components are polynomials is presented in [25].

In [12], 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[1] which is the combination of the discrete direct and the continuous inverse F-transform and derived an optimization of parameters of uniform 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,[2] the FT-smoothing filter provides very satisfactory estimates of PDFs. Moreover, similarly to the vector quantization based on Parzen windows, or the finite Gaussian mixture, the FT-smoothing filter decreases significantly the model complexity being proportional for Parzen windows to the number of sample data which can lead to the memory storage problem. A disadvantage of the FT-smoothing filter is an overfitting for a smaller bandwidth of basic functions which is among others caused by a limitation of the original definition of fuzzy partition.

In this paper, we generalize the FT-smoothing filter proposed in [12], which is based on the original conception of the fuzzy partition, to better check the smooth property of the resulted function. As we have mentioned above, the original definition of fuzzy partition has some limitations for smoothing 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 [30], where an extension of the fuzzy partition to a fuzzy r-partition (r is a natural

[1]Note that the abbreviation "FT" means "based on the fuzzy transform".

[2]Note 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 assumption on the shape and parameters of PDFs (for a survey, see [11, 28]).

number grater than 0) with $2r$ active basic functions[3] and discrete and continuous $\mathrm{F}^{(\mathrm{r})}$-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 in [30] (see also [31]) uses advanced forms of fuzzy numbers for basic functions modeling based on parametric shape functions and shows a smoothing effect on functions obtained from the $\mathrm{F}^{(\mathrm{r})}$-transform. Following the Stefanini's idea, we propose an extended fuzzy r-partition as a special case of the fuzzy r-cover, where we do not suppose that r is a natural number. 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 as 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 functions (see Remark 2.3).

To check the quality of the generalized FT-smoothing filter we verify its approximative behavior using the modulus of continuity, smoothing properties expressed by the total variation of smoothed function and statistical properties including the white noise reduction and the asymptotic properties of the filter as an estimator.[4] The approximation behavior using the modulus of continuity has been investigated in [22] and we use here only a slight reformulation of the modulus of continuity. An analysis of the smoothed functions from the total variation point of view has been presented in [30]. Note that in this paper the total variation of smoothed functions has a different form, which is required by the considered discrete design, and the results are presented graphically. Here, we prove a statement saying that under some conditions the total variation tends to decrease for higher values of r (see Corollary 4.7). The fruitfulness of the F-transform in many applications is closely related to the fact that the F-transform components satisfy the weighted least square criterion (see Theorem 3.1 in [24]). Here, one can recognize the derivation of the Nadaraya-Watson (NW) estimator in the kernel regression to find unknown function from a data sample.[5] This fact leads one to carry-out a statistical analysis of the FT-smoothing filter estimator analogously to the case of the NW estimator and to compare both results. Note that the NW estimator belongs among the traditional approaches to the kernel regression and it is a special case of local polynomial kernel estimators.[6] Here, we restrict our study of the statistical properties of the FT-smoothing

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

[4]The FT-smoothing filter as an estimator of an unknown function at a given point will be called the FT-smoothing filter estimator.

[5]For the definition, see (16) on page 13. Note that the NW estimator was independently introduced by Nadaraya [20] and Watson [35]. For the interested reader, we refer to [34] for an excellent introduction to the kernel smoothing or to [10, 11, 21, 28, 29].

[6]It is to be remarkable that the NW estimator is derived as the polynomial of the degree 0. There are many results on the estimators determined by the polynomial of the degree greater than 0, and theoretically, provide a better estimates (see [8, 21, 11]). Hence, an extension of the F-transform to higher order presented in [25] seems to be the right way how to improve the quality of smoothing.

filter estimator to the fixed design and express its bias (Bias) and variance (Var).[7] Further, we show that a white noise reduction can be ensured for the denser fuzzy r-partitions. Note that this result can be used in many practical applications.

As Stefanini has mentioned in [30], the $\mathrm{F}^{(\mathrm{r})}$-transform can be expressed in terms of the linear filters as a moving average operator. Hence, we obtain that the FT-smoothing filter estimator is asymptotically unbiased similarly to the local polynomial kernel estimators. Nevertheless, this fact does not say anything about the rate of convergence to compare different approaches and, moreover, the asymptotic bias and variance are used to derive the optimal value of bandwidth with respect to the asymptotic mean square error (see $h^{\mathrm{NW}}_{\mathrm{AMSE}}$ on p. 15). Here, we propose convergence conditions and derive formulas for the bias and variance of the estimator based on our filter. The results are compared with that obtained by the NW estimator. Note that, for this purpose, we use the equally spaced fixed design and fuzzy r-partitions are expressed in the terms of kernels. A main reason for such restriction and the kernel expression is a high complexity of the asymptotic expression of the covariance in (43) of Theorem 4.8 in the random design and the possibility to compare results of the FT-smoothing filter estimator and NW estimator, respectively.

The paper is structured as follows. The following section gives formal definitions of countable fuzzy r-cover and fuzzy r-partition as a special case of the fuzzy r-cover. Further, the concept of a (uniform) fuzzy r-partition determined by a set of nodes and a set of basal fuzzy sets is introduced. 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 including a stochastic version. A comparison of the F-transform component as an estimator of unknown function values and the NW estimator is performed and the asymptotic properties of the NW estimator in the fixed equally spaced design are recalled. The main part of this paper is the fourth section, where a generalization of the FT-smoothing filter is given. Finally, the mentioned properties as the asymptotic behavior or statistical properties are investigated. An illustrative example assuming financial data for a practical comparison of the FT-smoother and NW smoother is given in the fifth section. The last section is a conclusion.

2. Fuzzy r-partitions

We shall use $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ to denote the set of all natural, integer and real numbers, respectively. We shall use $\mathbb{I}$ to denote a (finite or denumerable) set of consecutive integers. Usually, the set $\mathbb{I}$ has one of the form $\mathbb{I} = \{1, \ldots, n\}$, or

[7] Let $\hat{g}(x)$ be an estimator of an unknown function $g(x)$. The *bias* and *variance* of the estimator $\hat{g}(x)$ of $g(x)$ is defined by $\mathrm{Bias}(\hat{g}(x)) = \mathrm{E}(\hat{g}(x) - g(x)) = \mathrm{E}(\hat{g}(x)) - g(x)$ and $\mathrm{Var}(\hat{g}(x)) = \mathrm{E}((\hat{g}(x) - g(x))^2)$, respectively, where E denotes the expected value over the sampling distribution of $\hat{g}(x)$.

$\mathbb{I} = \{-n, -n+1, \dots, n-1, n\}$ for some $n \in \mathbb{N}$ in the finite case, or $\mathbb{I} = \mathbb{Z}$ in the infinite case.

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

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

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 $\mathrm{Ker}(A) = \{x\}$ for a suitable $x \in \mathbb{R}$. In this paper, we shall suppose that each fuzzy set is continuous, convex and has the non-empty kernel.[8] Note that the choice of the fuzzy sets shapes is motivated by the shapes of fuzzy sets used in the original definition of fuzzy partition [24] and this choice seems to be natural. Finally, we denote $R = [a, b]$.

In order to define a generalization of fuzzy partition proposed in [24], let us start with more general concept called a fuzzy r-cover of a real interval R. A motivation of this step is closely connected with a solution to the problem of finding a better approximation of a function in the endpoints of the given interval. An extension of the fuzzy partition has been also used in [25].

Definition 2.1 (Fuzzy r-cover). Let R be a real interval and $r \geq 1$ be a real number. A *countable fuzzy r-cover of R* is a collection $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ of non-empty fuzzy sets that satisfies

$$\sum_{i \in \mathbb{I}} A_i(x) \geq r \tag{2}$$

for any $x \in R$.

Note that the condition (2) is motivated by a natural generalization of a countable cover of a set by sets, i.e., R is covered by sets from a collection $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$, if $R \subseteq \bigcup_{i \in \mathbb{I}} A_i$, whereas $A_i \cap A_j \neq \emptyset$ for $i \neq j$ is not supposed in general. A simple but very useful consequence of the definition of fuzzy r-cover is as follows.

Proposition 2.1. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be disjoint fuzzy r_1-cover and r_2-cover of a real interval 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.*

According to the cardinality of the collection of $\mathcal{A}$, we can distinguish two cases of fuzzy r-covers. We shall say that a fuzzy r-cover is *finite*, if it contains a finite number of fuzzy sets, and *infinite*, otherwise. If we deal with a finite fuzzy r-cover of a real interval R, then we shall usually write $\mathcal{A} = \{A_1, \dots, A_n\}$, or $\mathcal{A} = \{A_{i-n}, \dots, A_i, \dots, A_{i+n}\}$. Notice that the sum in the r-Ruspini condition

[8] Note that if $\mathrm{Supp}(A) = \mathrm{cl}\{x \in \mathbb{R} \mid A(x) > 0\}$, where cl is the closure operator, then the considered fuzzy sets are fuzzy intervals or, accepting the assumption of the normality, fuzzy numbers (see [13, 7]).

is an infinite sum for infinite fuzzy r-covers of R which could be seen as a trouble for a practical manipulation. However, for many infinite fuzzy r-partitions, the infinite sum may be successfully transformed to a finite sum (see Example 2.1).

Since we work with a more general concept of fuzzy cover, it is useful to know fuzzy sets giving a non-zero function value at a point x of R. Let $\mathcal{A}$ be a fuzzy r-cover of R. Then we define a mapping

$$d_{\mathcal{A}} : R \to 2^{\mathbb{Z}} \tag{3}$$

such that $A_i(x) > 0$ for any $i \in d_{\mathcal{A}}(x)$ and $A_i(x) = 0$ for any $i \in \mathbb{I} \setminus d_{\mathcal{A}}(x)$ for any $x \in R$. A number of the "active" fuzzy sets at a point x of R is given as a number of elements contained in the set $d_{\mathcal{A}}(x)$. One can check easily that if $|d_{\mathcal{A}}(x)| = \infty$, then $\mathcal{A}$ is an infinite fuzzy r-cover and if $\mathcal{A}$ is a finite fuzzy r-cover, then $|d_{\mathcal{A}}(x)| < \infty$. Inverse implications are not true as the following simple counter-example demonstrates.

Example 2.1. Let us consider $R = (0, 1]$ and define

$$A_n(x) = f(x, 2^{-n-1}, 2^{-n}, 2^{-n+1})$$

for any $n \in \mathbb{N}$, where

$$f(x, a, b, c) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x < b; \\ \frac{c-x}{c-b}, & b \le x \le c; \\ 0, & c < x. \end{cases}$$

One can see that $\min(A_n(x), A_{n+1}(x)) > 0$ if and only if $x \in (2^{-n-1}, 2^{-n})$. Since

$$A_n(x) + A_{n+1}(x) = \frac{x - \frac{1}{2^{n+1}}}{\frac{1}{2^n} - \frac{1}{2^{n+1}}} + \frac{\frac{1}{2^n} - x}{\frac{1}{2^n} - \frac{1}{2^{n+1}}} = 1$$

for any $x \in [2^{-n-1}, 2^{-n}]$, the set $\mathcal{A} = \{A_i \mid n \in \mathbb{N}\}$ forms an infinite fuzzy 1-cover of R. On the other hand, we have $|d_{\mathcal{A}}(x)| \le 2$ for any $x \in R$. $\square$

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

Definition 2.2 (Fuzzy r-partition). Let R be a real interval and $r \ge 1$ be a real value. A collection $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ of non-empty fuzzy sets is called a *countable fuzzy r-partition of R*, if there exists a fuzzy r-cover $\mathcal{B} = \{B_i \mid i \in \mathbb{I}\}$ of R such that

1. $A_i = B_i \restriction R$ for any $i \in \mathbb{I}$,
2. $\sum_{i \in \mathbb{I}} A_i(x) = r$ for any $x \in R$.

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

It is easy to see, if $r = 1$, then the condition 2 coincides with the Ruspini condition (see [27]) used in the original definition of the fuzzy partition (Definition 1 in [22]). This motivates us to call the condition as the *r-Ruspini condition*. Simply speaking, a fuzzy r-partition of R is the restriction of functions of a fuzzy r-cover to R, where, moreover, a requirement on fuzzy sets to be mutually disjoint is considered. In comparison with the original definition in [24], the basic functions in our conception are also defined over nodes that need not belong to R and more than one node can be contained in supports of basic functions in general. This trick enables us to regulate the smoothness of resulted functions better.

Similarly to the fuzzy r-cover denotation, we shall say that a fuzzy r-partition of A is *finite* or *infinite*, if the set of all basic functions is finite or infinite, respectively. In the case that $\mathcal{A}$ is a finite fuzzy r-partition, then we shall write $\mathcal{A} = \{A_1, \ldots, A_n\}$ or $\mathcal{A} = \{A_{i-n}, \ldots, A_i, \ldots, A_{i+n}\}$.

Remark 2.2. 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 [31], where each fuzzy r-partition can be derived as the combination of fuzzy 1-partitions.

It is easy to see that the collection of all restricted fuzzy sets to the interval $(0, 1]$ introduced in Example 2.1 forms an infinite fuzzy 1-partition of $(0, 1]$. Further examples will be mentioned later.

In this paper, we are interested in a special case of fuzzy r-partitions, namely, a *uniform fuzzy r-partition determined by a basal fuzzy sets*. A main reason for this step is the investigation of the asymptotic properties of the FT-smoothing filter estimator in a spirit of asymptotic properties of kernel smoothing estimators, although, many results hold true without this restriction. Note that the uniform fuzzy r-partition is also profitable from an optimization point of view (see comments in [12]).

Definition 2.3. A *basal fuzzy set* is a continuous, convex, normal fuzzy set $S : \mathbb{R} \to [0, 1]$ such that $S(x) = S(-x)$ for any $x \in \mathbb{R}$ and

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

Obviously, the condition $S(x) = S(-x)$ for any $x \in \mathbb{R}$ says that each basal fuzzy set is symmetric which implies

$$\int_{-\infty}^{\infty} x S(x) dx = 0. \tag{4}$$

Note that this equality enables us to significantly simplify the description of e.g. the Bias, Var, MSE, or AMSE, that will be investigated in the subsection devoted to the statistical analysis of the FT-smoothing filter. Let S be a basal fuzzy set and $h > 0$ be a real number. On can check easily that also

$$S_h(x) = S\left(\frac{x}{h}\right) \tag{5}$$

is a basal fuzzy set. The value h is called the *bandwidth* or *window width*.

Definition 2.4. Let R be a real interval, $\mathcal{T} = \{t_i \mid i \in \mathbb{I}\}$ be an ordered set of nodes from $\mathbb{R}$ with $t_i < t_{i+1}$ for any $i \in \mathbb{I}$ and $\mathcal{S} = \{S^{(i)} \mid i \in \mathbb{I}\}$ be a set of basal fuzzy sets. A *fuzzy r-partition of R determined by* $(\mathcal{T}, \mathcal{S})$ is a fuzzy r-partition $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ of R such that

$$A_i(x) = S^{(i)}(x - t_i) \tag{6}$$

holds for any $x \in R$ and $i \in \mathbb{I}$. We shall say that a fuzzy r-partition of R determined by $(\mathcal{T}, \mathcal{S})$ is *uniform*, if $S = S^{(i)}$ for any $i \in \mathbb{I}$ 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 $i \in \mathbb{I}$, we shall also write $(\mathcal{T}, S)$ instead of $(\mathcal{T}, \mathcal{S})$. It is easy to see that $d_{\mathcal{A}}(x)$ for the finite uniform fuzzy r-partitions has a form $\{i_0, i_0 + 1, \dots, i_0 + n\}$ for a suitable $i_0 \in \mathbb{Z}$. Note that $t_i \notin R$ for some $i \in \mathbb{I}$, in general, however, $S^{(i)}(x - t_i) > 0$ for some $x \in R$.

Proposition 2.2. *For any finite uniform fuzzy r-partitions $\mathcal{A}$ of R and for any $x, y \in R$, we have $||d_{\mathcal{A}}(x)| - |d_{\mathcal{A}}(y)|| \leq 1$.*

Proof. Let $\mathcal{A}$ be a finite uniform fuzzy r-partitions of R and $x, y \in R$. Without loss of generality, let us suppose that $S(z) = 0$ for all $z \notin (-1, 1)$ and $d_{\mathcal{A}}(x) = \{i_0, i_0 + 1, \dots, i_0 + n\}$ and $d_{\mathcal{A}}(y) = \{j_0, j_0 + 1, \dots, j_0 + m\}$. Put $x_i = t_{i_0+i} - x$, where $i = 0, \dots, n$, and $y_j = t_{j_0+i} - y$, where $j = 0, \dots, m$. It is easy to see that $x_i = x_1 + iu$ and $y_j = y_1 + ju$, where $u = t_{i+1} - t_i$ is the constant derived from the uniformity of fuzzy r-partition. Moreover, we have $-1 < x_0 \leq y_0 \leq -1 + u$, or $-1 < y_0 < x_0 \leq -1 + u$, otherwise, $t_{i_0} \notin d_{\mathcal{A}}(x)$ or $t_{j_0} \notin d_{\mathcal{A}}(y)$.

Here, let us suppose that $x_0 \leq y_0$. Analogously, one can prove the statement for the second case. Notice that $1 - u \leq x_n \leq y_m < 1$. A simple consequence of the construction of points x_i and y_j is the fact that $y_n \in [1 - u, 1)$ or $y_n \notin [1 - u, 1)$. In the first case, we necessary obtain $m = n$ and, hence, $||d_{\mathcal{A}}(x)| - |d_{\mathcal{A}}(y)|| = 0$. In the second one, we have $m = n - 1$ (recall that $y_n - y_{n-1} = u$) and thus $||d_{\mathcal{A}}(x)| - |d_{\mathcal{A}}(y)|| = 1$. Hence, we may write $||d_{\mathcal{A}}(x)| - |d_{\mathcal{A}}(y)|| \leq 1$ in general. □

Remark 2.3. It is not easy to show something about the structure of the uniform fuzzy r-partition in a general setting. To demonstrate the difficulty of such investigation, we prove the claim: *Let $\mathcal{A}$ be a finite uniform fuzzy r-partitions of R determined by $(\mathcal{T}, S_h)$ (put $u = t_{i+1} - t_i$) such that r is a natural number and there is a real number a with $0 < a < 1$ and $S_h(su) = 1 - sa$ for any $s = 0, 1, 2, \dots$ Then $ru = h$.*

In fact, one can see from the assumption on the basal fuzzy set S_h and the uniformity of the fuzzy r-partition $\mathcal{A}$ that

$$\begin{gathered}
S_h(t_i - t_i) = S_h(0) = 1 \\
S_h(t_i - t_{i+1}) = S_h(t_i - t_{i-1}) = 1 - a \\
\vdots \\
S_h(t_i - t_{i+s-1}) = S_h(t_i - t_{i-s-1}) = 1 - (s-1)a,
\end{gathered}$$

where s is the least natural number for which $sa \geq 1$. One can check that $s \geq 2$. We shall prove that $sa = 1$ which gives that $S_h(x - t_{i-s}) > 0$ and $S_h(y - t_{i+s}) > 0$ for any $x \in [t_{i-s}, t_i[$ and $y \in]t_i, t_{i+s}]$, respectively, and $S_h(t_i - t_{i-s}) = S_h(t_i - t_{i+s}) = 0$. A simple consequence of this fact is $h = t_{i+s} - t_i = s(t_{i+1} - t_i) = su$. From the r-Ruspini condition, we have (note that under our assumption on r we have $0 < (s-1)a < 1$):

$$\begin{gathered}
r = 1 + 2(1-a) + 2(1-2a) + \cdots + 2(1-(s-1)a) = \\
1 + 2(s-1) - a(s-1)s = \\
1 + (s-1) + (s-1) - a(s-1)s = s + (1-as)(s-1).
\end{gathered}$$

Let us suppose that $sa > 1$. It is easy to show that $-1 < 1 - sa < 0$. Since s is a natural number and $s \geq 2$, then $s + (1-sa)(s-1)$ cannot be a natural number which is a contradiction with the assumption on r. Hence, we obtain $sa = 1$ and $r = s$, i.e. $ru = h$.

Analogous statements could be very interesting for the development of the theory about the fuzzy r-partitions, but their realization is rather complex and out of the scope of this paper. On the other hand, from the application perspective, the user can require some special conditions on the (uniform) fuzzy r-partitions as, for example, $ru = h$ (see Examples 2.4 and 2.5).

The following examples of uniform fuzzy r-partitions are based on the basic functions that have been used in [24]. For an interesting example of a non-uniform fuzzy r-partition based on the parameterized fuzzy numbers (see [31]), we refer to [30].

Example 2.4 (Triangle fuzzy r-partition). The triangle basal fuzzy sets is given by

$$S_h(x) = \max\left(0, \frac{h - |x|}{h}\right) \tag{7}$$

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 $[0, 10]$ with the bandwidth $h = 2$ and a uniform fuzzy 9-partition of $[0, 10]$ with the bandwidth $h = 6$. For a simpler orientation, we split the uniform fuzzy r-partitions onto two parts pictured by normal and dashed lines that are again fuzzy partitions.

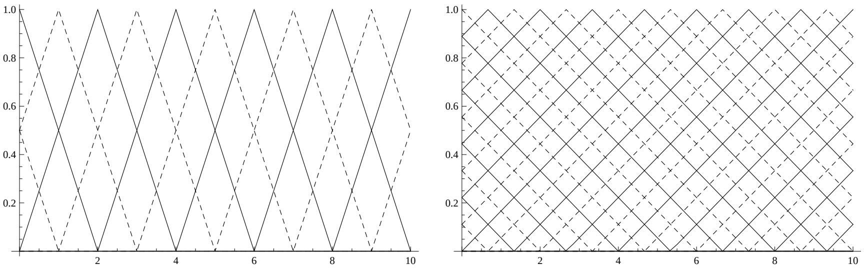

Figure 1: Triangle uniform fuzzy 2-partition for $h = 2$ (left) and 9-partition for $h = 6$ (right).

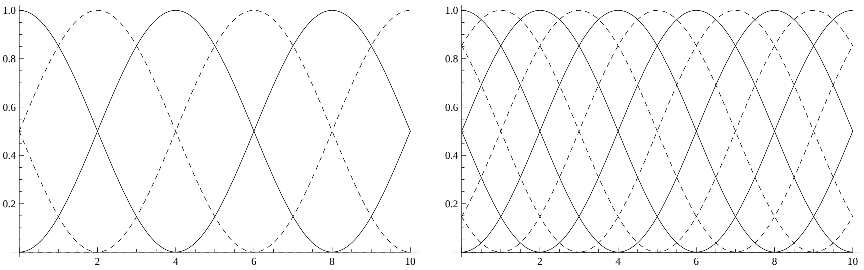

Figure 2: Raised cosine uniform fuzzy 2-partition (left) and 4-partition (right) for $h = 4$.

Example 2.5 (Raised Cosine fuzzy r-partition). The raised cosine basal fuzzy set is given by

$$S_h(x) = \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{x}{h}\pi\right)\right), & x \in [-h, h]; \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

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

Remark 2.6. One can notice that when fixing a bandwidth h a greater value of r requires a greater number of basic functions and the fuzzy r-partition is denser, i.e., the number of nodes t_i is greater. It is easy to see that fixing the value of r a smaller bandwidth leads to a denser fuzzy r-partition and vice versa.

As we have mentioned in Introduction, the components of the F-transform are derived analogously to the values of the NW estimator. To show a relation between both notions, let us establish the concept of kernel (see e.g. [9, 11, 21, 28, 34]). A *kernel* is a continuous, symmetric, unimodal function $K : \mathbb{R} \to [0, \infty)$ having the following properties:

(i) $\int_{-\infty}^{\infty} K(x)dx = 1$,

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

Obviously, each kernel is a symmetric density function with a single mode. Typical examples of kernels are the Uniform, Triangular, Cosine, Epanechnikov or Gaussian density functions (see e.g. [9, 11, 29]). Let $h > 0$ be a real number and K be a kernel, then $K_h(x) = 1/hK(x/h)$ is again a kernel, where h is the bandwidth of K (see [34]). For our investigation, the following relation between basal fuzzy sets and kernels is fundamental.

Proposition 2.3. *There is a one-to-one correspondence between the classes of all basal fuzzy sets and all kernels.*

Proof. Define

$$F(S)(x) = \frac{S(x)}{\int_{-\infty}^{+\infty} S(x)dx} \quad \text{and} \quad G(K)(x) = \frac{K(x)}{K(0)}$$

for any basal fuzzy set S and any kernel K. It is easy to see that $F(S)$ is a kernel and $G(K)$ is a basal fuzzy set. One can simply check that $F \circ G(K) = K$ and $G \circ F(S) = S$. □

A valuable consequence of this proposition is the fact that the well known results for the kernels and kernel smoothers can be adopted for the basal fuzzy sets and potentially for an FT-smoothing filter estimator based on the F-transform.

3. Direct discrete fuzzy transform

Let us introduce the concept of the (direct) discrete fuzzy transform which assigns, using basic functions of fuzzy r-partitions, to a finite real function g a vector of real numbers representing this function g. We shall use $\mathrm{Dom}(g)$ to denote the domain of a function g.

Let R be a real interval and $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ be a fuzzy r-partition of R determined by $(\mathcal{T}, \mathcal{S})$ and $s \in \mathbb{N}$. We shall 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 shall say that X is *sufficiently dense with respect to* $\mathcal{A}$.[9] Note that if a set $\{x_j \mid j = 1, \dots, k\}$ of reals is sufficiently dense with respect to a fuzzy r-partition $\mathcal{A}$ and the domains of basic functions of $\mathcal{A}$ are bounded, i.e. $\mathrm{Dom}(A_i) \subseteq [a_i, b_i]$ for any $i \in \mathbb{I}$, then $\mathcal{A}$ is a finite fuzzy r-partition.

[9] Usually, it is sufficient to suppose that sets of nodes are 1-dense. Nevertheless, there are some cases when the assumption of s-density for $s > 1$ leads to a better result (see Corollary 4.10).

Definition 3.1 ([24]). Let R be a real interval, g be a finite real function given at the nodes $x_1 < \cdots < x_n$ with $\mathrm{Dom}(g) \subseteq R$ and $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ be a fuzzy r-partition of R determined by $(\mathcal{T}, \mathcal{S})$ such that $\mathrm{Dom}(g)$ is sufficiently dense with respect to $\mathcal{A}$. We shall say that a collection of real numbers $\{F_i \mid i \in \mathbb{I}\}$ is the *discrete fuzzy (F-)transform of g with respect to $\mathcal{A}$*, if

$$F_i = \frac{\sum_{j=1}^{n} g(x_j) A_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)}. \tag{9}$$

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

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

Notice that the assumption of "being sufficiently dense" used in the previous definition ensures the correctness of the formula (9), i.e. the denominator is different from 0 for each basic function A_i. Since A_i are determined by basal fuzzy sets $S^{(i)}$, then we may also write

$$F_i = \frac{\sum_{j=1}^{n} g(x_j) S^{(i)}(x_j - t_i)}{\sum_{j=1}^{n} S^{(i)}(x_j - t_i)}, \tag{10}$$

or even

$$F_i = \frac{\sum_{j=1}^{n} g(x_j) S(x_j - t_i)}{\sum_{j=1}^{n} S(x_j - t_i)}, \tag{11}$$

if a uniform fuzzy r-partition is supposed.

In [24], a very important statement mentioned below is presented. This proposition shows that the components of F-transform satisfy the weighted least square criterion.

Proposition 3.1. *Let R be a real interval, g be a real function given at the nodes $x_1 < \cdots < x_n$ with $\mathrm{Dom}(g) \subseteq R$ and $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ be a fuzzy r-partition of R. Then F_i minimizes the weighted least square criterion*

$$\Psi_i(y) = \sum_{j=1}^{n} (g(x_j) - y)^2 A_i(x_j). \tag{12}$$

The rest of this section is devoted to a statistical analysis of the F-transform components based on a uniform fuzzy r-partition from the kernel-based non-parametric regression perspective.[10] More precisely, we are interested in the fixed equally spaced design context.

[10] It is possible to deal with non-uniform fuzzy r-partitions, but the results are more complicated and non-transparent.

Before commencing our study, let us give some relevant terminology and notation. A *fixed design* consists of $x_1, \ldots, x_n$ 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. Sometimes, a fixed equally spaced design has a form $x_j = \frac{j}{n}$ for $j = 0, \ldots, n$. For the fixed design the response variables are assumed to satisfy

$$Y_j = g(x_j) + \varepsilon_j, \quad j = 1, \ldots, n, \tag{13}$$

where g is a (non-random) function and ε_j is a random variable representing the error in x_j having the following expected value, variance and covariance

$$\mathrm{E}(\varepsilon_j) = 0,\ \mathrm{Var}(\varepsilon_j) < \infty \text{ and } \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \tag{14}$$

for $i \neq j$, respectively. In our investigation, we restrict ourselves to the common choice $\mathrm{Var}(\varepsilon_j) = \sigma^2$ for any $j = 1, \ldots, n$ (a homoscedastic model). One can prove easily that $\mathrm{E}(Y_j) = g(x_j)$, $\mathrm{Var}(Y_j) = \mathrm{Var}(\varepsilon_j) = \sigma^2$ and $\mathrm{Cov}(Y_i, Y_j) = 0$ for any $i \neq j$.[11]

Let R be a real interval, $(x_1, Y_1), \ldots, (x_n, Y_n)$ define a finite random function, where $Y_1 \ldots, Y_n$ are given by (13) and $X = \{x_1, \ldots, x_n\} \subseteq R$, and $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ be a uniform fuzzy r-partition of R determined by $(\mathcal{T}, S)$ such that X is sufficiently dense with respect to $\mathcal{A}$. We shall say that a collection of random variables $\{\Phi_i \mid i \in \mathbb{I}\}$ is the *discrete stochastic F-transform of the finite random function defined by* $(x_1, Y_1), \ldots, (x_n, Y_n)$ *with respect to* $\mathcal{A}$, if

$$\Phi_i = \frac{\sum_{j=1}^{n} Y_j S(x_j - t_i)}{\sum_{j=1}^{n} S(x_j - t_i)}. \tag{15}$$

Note that if $\{\Phi_i \mid i \in \mathbb{I}\}$ are the components of the stochastic F-transform of a finite random function defined by $(x_1, Y_1), \ldots, (x_n, Y_n)$, where $Y_j = g(x_j) + \varepsilon_j$ for any $j = 1, \ldots, n$, and $\{F_i \mid i \in \mathbb{I}\}$ are the F-transform components of the (non-random) finite function g, then the expected value of the random variable Φ_i is equal to F_i (i.e. $\mathrm{E}(\Phi_i) = F_i$) for any $i \in \mathbb{I}$.

Let us recall that the aim of the kernel-based nonparametric regression is to estimate the unknown function g. There are many methods based on one or more than one kernels how to find a "good" estimation of g (see e.g. [9, 10, 11, 21, 28, 29, 34]). Here, we restrict ourselves to one of them called the Nadaraya-Watson (NW) estimator:

$$\hat{g}_{\mathrm{NW}}(x) = \frac{\sum_{j=1}^{n} Y_j K(x_j - x)}{\sum_{j=1}^{n} K(x_j - x)}, \tag{16}$$

where Y_j is expressed by (13) and K is a kernel. Comparing the NW estimator with the formula (15) for the stochastic F-transform components over a random function given by $(x_1, Y_1), \ldots, (x_n, Y_n)$, one can see the similarity. More precisely, we may state the following proposition.

[11] For example, $\mathrm{Cov}(Y_i, Y_j) = \mathrm{E}((Y_i - g(x_i)(Y_j - g(x_j))) = \mathrm{E}((\varepsilon_i - 0)(\varepsilon_j - 0)) = \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$

Proposition 3.2. *Let R be a real interval and $\mathcal{A} = \{A_i \mid i \in \mathbb{I}\}$ be a uniform fuzzy r-partition of R determined by $(\mathcal{T}, S)$. If $Y_1, \ldots, Y_n$ are random variables defined by (13) and $\hat{g}_{\mathrm{NW}}$ is the NW estimator with*

$$K(x) = \frac{S(x)}{\int_{-\infty}^{+\infty} S(x)dx},$$

then $\Phi_i = \hat{g}_{\mathrm{NW}}(t_i)$ for any $t_i \in \mathcal{T}$.

PROOF. This is a straightforward consequence of Proposition 2.3 and the definition of Φ_i. $\square$

Since the components Φ_i coincide with the estimates $\hat{g}_{\mathrm{NW}}(t_i)$ of function values $g(t_i)$, we may use the results of the NW estimator to investigate statistical properties of our smoothing filter as an estimator.[12] Recall several facts about the asymptotic behavior of the NW estimator which are among others used to compare the quality of kernel smoothers and to derive an optimal bandwidth for the "best smoothing".

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

$$Y_j = g(x_j) + \varepsilon_j, \quad j = 1, \ldots, n, \tag{17}$$

where $x_j = a + j\left(\frac{b-a}{n}\right)$ and $\mathrm{Var}(\varepsilon_j) = \sigma^2$. 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 assumptions in our analysis (cf. [34]):

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

(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 $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

(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 the interval $[a, b]$ in the fixed equally spaced design context is considered to be the unit interval. Recall that $\mathrm{Bias}(\hat{g}_{\mathrm{NW}}(x)) = \mathrm{E}(\hat{g}_{\mathrm{NW}}(x)) - g(x)$.

[12] Note that NW estimator is the simplest non-parametric estimator and can be improved by, for example, a linear local regression (see e.g. [9, 11, 21, 34]).

Theorem 3.3. *Under the assumptions (A1)-(A4), we may write*

$$\mathrm{Bias}(\hat{g}_{\mathrm{NW}}(x)) = \frac{g''(x)h^2\mu_2(K)}{2} + O\left(\frac{1}{n}\right) + o(h^2), \tag{18}$$

$$\mathrm{Var}(\hat{g}_{\mathrm{NW}}(x)) = \frac{\sigma^2(b-a)}{nh}R(K) + o\left(\frac{1}{nh}\right), \tag{19}$$

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

For the proof, we refer to e.g. [8, 9, 34]. Note that if one assumes the unit interval as $[a, b]$, then (19) can be rewritten as

$$\mathrm{Var}(\hat{g}_{\mathrm{NW}}(x)) = \frac{\sigma^2}{nh}R(K) + o\left(\frac{1}{nh}\right). \tag{20}$$

Further, we can see that $\hat{g}_{\mathrm{NW}}(x)$ is an asymptotically unbiased estimator of g. A simple consequence of this theorem is a computation of the *mean square error*

$$\begin{gathered}\mathrm{MSE}(\hat{g}_{\mathrm{NW}}(x)) = \mathrm{Var}(\hat{g}_{\mathrm{NW}}(x)) + \mathrm{Bias}(\hat{g}_{\mathrm{NW}}(x))^2 = \\ \frac{\sigma^2(b-a)}{nh}R(K) + o\left(\frac{1}{nh}\right) + \left(\frac{g''(x)h^2\mu_2(K)}{2} + O\left(\frac{1}{n}\right) + o(h^2)\right)^2 = \\ \frac{\sigma^2(b-a)}{nh}R(K) + \frac{(g''(x))^2h^4\mu_2(K)^2}{4} + o\left(\frac{1}{nh} + h^4\right).\end{gathered}$$

Let us denote

$$\mathrm{AMSE}(\hat{g}_{\mathrm{NW}}(x)) = \frac{\sigma^2(b-a)}{nh}R(K) + \frac{(g''(x))^2h^4\mu_2(K)^2}{4} \tag{21}$$

as the asymptotic MSE. The optimal value of the bandwidth h can be derived by setting the derivative of AMSE with respect to h equal to zero (for details, we refer to [34, 9, 8]). By a simple calculation we obtain

$$h^{\mathrm{NW}}_{\mathrm{AMSE}} = \left(\frac{\sigma^2(b-a)}{ng''(x)}C(K)\right)^{\frac{1}{5}} \tag{22}$$

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

4. FT-smoothing filter

In this section, we generalize the FT-smoothing filter based on the direct discrete F-transform and the inverse continuous F-transform introduced in [12]. Further, we investigate approximation, total variation and statistical properties of this smoothing filter.

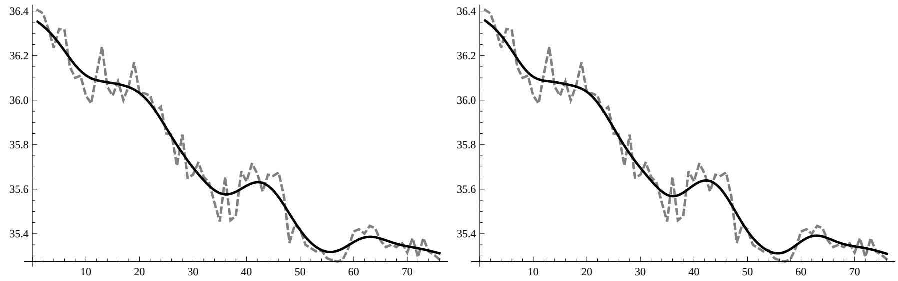

Figure 3: Application of the FT-smoothing filter on a geometric Brownian motion – triangle (left) and raised cosine (right) uniform fuzzy 5-partitions with $h = 5$.

4.1. Definition

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

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

$$\mathcal{F}_{\mathcal{A}}(g)(x) = \frac{1}{r} \sum_{i \in \mathbb{I}} F_i A_i(x) \tag{23}$$

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

Note that the formula in (23) has been used in [30] for the inverse discrete and continuous $\mathrm{F}^{(\mathrm{r})}$-transform. One can check easily that the linear combination of continuous functions is a continuous function. Hence, our definition is correct and $\mathcal{F}_{\mathcal{A}}$ is really a mapping to the set of all continuous functions on R. Comparing with the original approach to the inverse F-transform the proposed definition of FT-smoothing filter is only a slight modification of the inverse F-transform in the continuous design to obtain a linear combination for all elements from R. On Fig. 3, one can see an application of the FT-smoothing filter in smoothing a geometric Brownian motion in finance. In this way, a better imagination about the structure of this process can be obtained. We use triangle (left) and raised cosine (right) uniform fuzzy 5-partitions with the bandwidth $h = 5$. Obviously, there is no significant difference between the choice of the basal functions. Therefore, for the demonstrations in this paper, we restrict ourselves to the triangle uniform fuzzy r-partitions.

Recall that in the discrete design the resulted function of the inverse F-transform is again a finite function. More precisely, if g is a finite function with

$\mathrm{Dom}(g) = \{x_1, \ldots, x_n\}$, then

$$\hat{g}(x_i) = \mathcal{F}_{\mathcal{A}}(g)(x_i), \tag{24}$$

$i = 1, \ldots, n$, defines a finite function which is an approximation of the original function g (cf. Definition 5 in [24]). The following subsections are devoted to some properties of the FT-smoothing filter. To avoid some technical complications, we restrict ourselves to the finite fuzzy r-partition determined by $(\mathcal{T}, \mathcal{S})$. Nevertheless, most of the results are also true for the infinite fuzzy r-partition, and even for the fuzzy r-partitions that are not determined by basal fuzzy sets.

4.2. Basic properties of FT-smoothing filters

Let us suppose that $\mathcal{A} = \{A_1, \ldots, A_k\}$ be a fixed fuzzy r-partition of a real interval R determined by $(\mathcal{T}, \mathcal{S})$. Let us define the *partial addition* on $\mathrm{F}(R, \mathcal{A})$ (or on $\mathrm{CF}(R)$) by

$$(f + g)(x) = f(x) + g(x), \tag{25}$$

for any $f, g \in \mathrm{F}(R, \mathcal{A})$ such that $\mathrm{Dom}(f) = \mathrm{Dom}(g)$ and the *multiplication by a real number* on $\mathrm{F}(R, \mathcal{A})$ (or on $\mathrm{CF}(R)$) by

$$(af)(x) = af(x) \tag{26}$$

for any $f \in \mathrm{F}(R, \mathcal{A})$ and $a \in \mathbb{R}$. Let $f, g \in \mathrm{F}(R, \mathcal{A})$ (or $f, g \in \mathrm{CF}(R)$). We shall say that f is *less than or equal to* g and write $f \leq g$, if $\mathrm{Dom}(f) = \mathrm{Dom}(g)$ and $f(x) \leq g(x)$ for any $x \in \mathrm{Dom}(f)$. Obviously, the relation $\leq$ is a *partial ordering* on $\mathrm{F}(R, \mathcal{A})$ (or on $\mathrm{CF}(R)$).

Proposition 4.1. *Let* $f, g \in \mathrm{F}(R, \mathcal{A})$ *such that* $\mathrm{Dom}(f) = \mathrm{Dom}(g)$ *and* $a, b \in \mathbb{R}$. *Then*

$$\mathcal{F}_{\mathcal{A}}(af + bg) = a\mathcal{F}_{\mathcal{A}}(f) + b\mathcal{F}_{\mathcal{A}}(g). \tag{27}$$

If $f \leq g$ *in* $\mathrm{F}(R, \mathcal{A})$, *then* $\mathcal{F}_{\mathcal{A}}(f) \leq \mathcal{F}_{\mathcal{A}}(g)$ *in* $\mathrm{CF}(R)$.

Proof. Let $f, g \in \mathrm{F}(R, \mathcal{A})$ with $\mathrm{Dom}(f) = \mathrm{Dom}(g) = \{x_1, \ldots, x_n\}$ and $a, b \in \mathbb{R}$. Then

$$H_i = \frac{\sum_{j=1}^{n} (af + bg)(x_j) A_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)} =$$

$$\frac{a \sum_{j=1}^{n} f(x_j) A_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)} + \frac{b \sum_{j=1}^{n} g(x_j) A_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)} = aF_i + bG_i$$

holds for any $i = 1, \ldots, k$ and hence

$$\mathcal{F}_{\mathcal{A}}(af + bg)(x) = \frac{1}{r} \sum_{i=1}^{k} H_i A_i(x) = \frac{1}{r} \sum_{i=1}^{k} (aF_i + bG_i) A_i(x) =$$

$$\frac{a}{r} \sum_{i=1}^{k} F_i A_i(x) + \frac{b}{r} \sum_{i=1}^{k} G_i A_i(x) = a\mathcal{F}_{\mathcal{A}}(f) + b\mathcal{F}_{\mathcal{A}}(g).$$

If $f \leq g$ in $\mathrm{F}(R, \mathcal{A})$, then one can verify easily that $F_i \leq G_i$ for any $i = 1, \ldots, k$ and, hence, we obtain $\mathcal{F}_{\mathcal{A}}(f) \leq \mathcal{F}_{\mathcal{A}}(g)$. □

Proposition 4.2. *Let $g \in \mathrm{F}(R, \mathcal{A})$, $g(x) = c$ for any $x \in \mathrm{Dom}(g)$. Then $\mathcal{F}_{\mathcal{A}}(g)(x) = c$ for any $x \in R$.*

Proof. Let $g \in \mathrm{F}(R, \mathcal{A})$ and $g(x) = c$ for any $x \in \mathrm{Dom}(g) = \{x_1, \ldots, x_n\}$. Then

$$F_i = \frac{\sum_{j=1}^{n} cA_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)} = \frac{c\sum_{j=1}^{n} A_i(x_j)}{\sum_{j=1}^{n} A_i(x_j)} = c$$

for any $i = 1, \ldots, k$. Since $\frac{1}{r}\sum_{i=1}^{k} A_i(x) = 1$ for any $x \in R$, then

$$\mathcal{F}_{\mathcal{A}}(g)(x) = \frac{1}{r}\sum_{i=1}^{k} F_i A_i(x) = \frac{1}{r}\sum_{i=1}^{k} cA_i(x) = c\frac{1}{r}\sum_{k=1}^{k} A_k(x) = c$$

for any $x \in R$. □

4.3. Approximation behavior of FT-smoothing filter

To investigate an approximation behavior of FT-smoothing filters we shall define a parameterized modulus of continuity (cf. [14]). Let τ be a topology on R. A mapping $\delta : R \to \tau$ is called a *neighborhood function on* R, if $x \in \delta(x)$ holds for any $x \in R$. For example, if $R = [a, b]$ and τ is determined by the closed intervals on R, then a *neighborhood function on* R can be defined as $\delta(x) = [x - 2h, x + 2h] \cap R$, where h is the bandwidth of a basal fuzzy set. Intuitively, the mapping $\delta(x)$ should express some interesting neighborhood of the point x. Now, a parameterized modulus of continuity determined by δ may be introduced as follows.

Definition 4.2. Let g be a real function such that $\mathrm{Dom}(g) \subseteq R$, τ be a topology on R and $\delta : R \to \tau$ be a neighborhood mapping on R defined above. A *parameterized modulus of continuity* of the function g determined by δ is a mapping $\omega_\delta(g, \cdot) : R \to \mathbb{R}$ given by

$$\omega_\delta(g, x) = \sup_{r,s \in \delta(x) \cap \mathrm{Dom}(g)} |g(r) - g(s)|. \tag{28}$$

Obviously, $\omega_\delta(g, x)$ is the value saying how close the function values of g are at the points laying inside the neighborhood of the point x expressed by $\delta(t)$. Hence, smaller values of $\omega_\delta(g, x)$ show that the function g is smoother or continuous. If we want to have one number which characterizes the smoothness of a function g, then, analogously to the common modulus of continuity, we can consider

$$\omega_\delta(g) = \sup_{x \in R} \omega_\delta(g, x). \tag{29}$$

One can check easily that the common definition of modulus of continuity may be obtained by putting $\delta(x) = [x-h, x+h] \cap R$ for some $h > 0$. Obviously, if g is a continuous function on R, then $h \to 0$ implies $\omega_\delta(g) \to 0$, where $\delta(x) = [x-h, x+h] \cap R$. Let $A_i : \mathbb{R} \to [0, \infty)$ be a fuzzy set. We shall denote $\mathrm{Supp}(A_i) = \{x \in \mathbb{R} \mid A_i(x) > 0\}$ the *support* of the fuzzy set A_i.

Theorem 4.3. *Let $g \in \mathrm{F}(R, \mathcal{A})$. Then*

$$|g(x) - \mathcal{F}_\mathcal{A}(g)(x)| \leq \omega_\delta(g, x), \tag{30}$$

holds for any $x \in \mathrm{Dom}(g)$, where

$$\delta(x) = \bigcup_{i \in d_\mathcal{A}(x)} \mathrm{Supp}(A_i). \tag{31}$$

PROOF. Let $x \in \mathrm{Dom}(g)$ and $i \in d_\mathcal{A}(x)$. Since $A_i(x_j) = 0$ for any $x_j \in \mathrm{Dom}(g)$ such that $x_j \notin \delta(x)$, then

$$|g(x) - F_i| \leq \left| g(x) - \frac{\sum_{j=1}^n g(x_j)A_i(x_j)}{\sum_{j=1}^n A_i(x_j)} \right| = \left| \frac{\sum_{j=1}^n (g(x) - g(x_j))A_i(x_j)}{\sum_{j=1}^n A_i(x_j)} \right|$$
$$\leq \frac{\sum_{j=1}^n |g(x) - g(x_j)|A_i(x_j)}{\sum_{j=1}^n A_i(x_j)} = \frac{\sum_{x_j \in \delta(x)} |g(x) - g(x_j)|A_i(x_j)}{\sum_{x_j \in \delta(x)} A_i(x_j)} \leq$$
$$\frac{\sum_{x_j \in \delta(x)} \omega_\delta(g, x)A_i(x_j)}{\sum_{x_j \in \delta(x)} A_i(x_j)} = \omega_\delta(g, x).$$

Hence, we obtain

$$|g(x) - \mathcal{F}_\mathcal{A}(g)(x)| = \left| g(x) - \frac{1}{r}\sum_{i=1}^k F_i A_i(x) \right| =$$
$$\left| \frac{1}{r} \sum_{i \in d_\mathcal{A}(x)} (g(x) - F_i)A_i(x) \right| \leq$$
$$\frac{1}{r} \sum_{i \in d_\mathcal{A}(x)} |g(x) - F_i|A_i(x) \leq \frac{1}{r} \sum_{i \in d_\mathcal{A}(x)} \omega_\delta(g, x)A_i(x) = \omega_\delta(g, x)$$

and the proof is finished. □

Corollary 4.4. *Let $g \in \mathrm{F}(R, \mathcal{A})$. Then*

$$|g(x) - \mathcal{F}_\mathcal{A}(g)(x)| \leq \omega_\delta(g), \tag{32}$$

holds for any $x \in \mathrm{Dom}(g)$, where δ is defined by (31).

According to Remark 2.6, for a fixed bandwidth h and a higher value of r there is a need to use a higher number of nodes t_i (and vice versa) which generally implies a higher number of active basic functions for x described by $d_{\mathcal{A}}(x)$. Hence, the interval $\delta(x)$ defined by (31) has a greater width. A simple consequence of the definition of the modulus of continuity and the inequality in (30) is the fact that the distance between $g(x)$ and $\mathcal{F}_{\mathcal{A}}(g)(x)$ may be greater for higher values of r (and vice versa). An explanation of this performance is that the higher values of r (for a fix bandwidth h) tend to more smoothed functions and thus to greater distances from the original values $g(x)$ (see the following subsection). From Fig. 4, one can see, however, that the influence of the size of

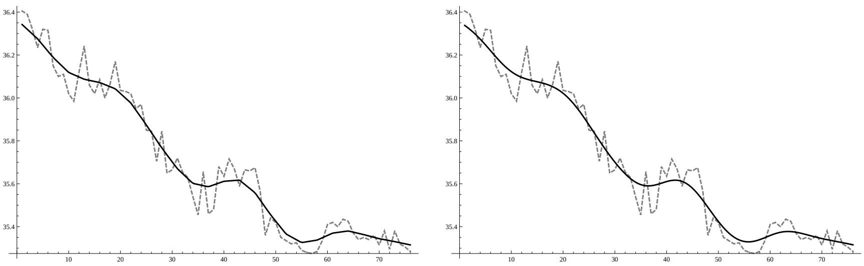

Figure 4: Approximation of a geometric Brownian motion for $r = 2$ (left) and $r = 18$ (right) with the same $h = 6$.

the value of r with the same bandwidth h on the approximation of a function is not too significant. Only the resulted function seems to be a bit more smoothed.

The effect for a fixed value of r can be derived similarly. In this case, lower values of h tend to give smaller width of $\delta(x)$ in general even if $d_{\mathcal{A}}(x)$ contains more indexes. In fact, we can generally write $\delta(x) \subseteq [x - 2h, x + 2h]$. Hence, lower values of h give a smaller distance between $g(x)$ and $\mathcal{F}_{\mathcal{A}}(g)(x)$ and the FT-smoothing filter better approximates the original values $g(x)$ as can be seen on Fig 5. Clearly, nobody is surprised by such performance of the FT-smoothing filter, because the major influence on the approximation of a function has the bandwidth h. The same statement is also true for other kernel smoothing filters.

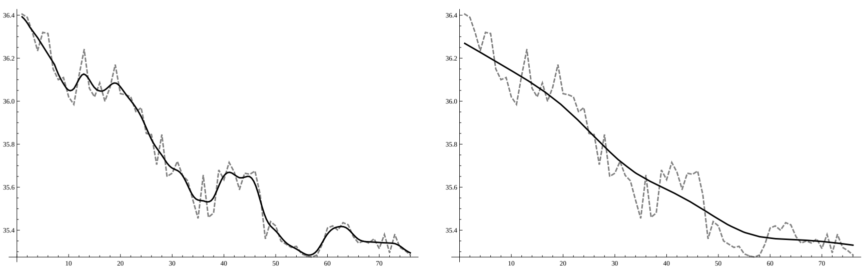

Figure 5: Approximation of a geometric Brownian motion for $h = 2$ (left) and $h = 12$ (right) with the same $r = 4$.

Supposing a continuous function g on R that is known only at nodes $x_1 < \cdots < x_n$ for some large number n, then obviously $\mathcal{F}_{\mathcal{A}}(g)(x)$ gives a very good approximation of g for small values of h. Note that this fact has been also pointed in the seminal paper [24].

4.4. Total variation of smoothed functions by FT-smoothing filter

In [30], Stefanini presented the influence of the values of r to the smoothness of resulted functions. For this purpose, the total variation is used (see e.g. [36]) and the results are graphicly presented. In this part, we prove a statement about the total variation of smoothed functions by the FT-smoothing filter.

Let us recall the definition of total variation of a real-valued function. A partition of a real interval $R = [a, b]$ is a finite order set $P = \{a = x_1 \leq \cdots \leq x_n = b\}$. Let $\mathcal{P}_R$ be the collections of all partitions of R. For a function g on R, the *total variation of g on R corresponding to a partition $P \in \mathcal{P}_R$* is defined by

$$V_R(g, P) = \sum_{j=1}^{n-1} |g(x_{j+1}) - g(x_j)|. \tag{33}$$

A *total variation of a function g on R* is defined by

$$V_R(g) = \sup_{P \in \mathcal{P}_R} V_R(g, P). \tag{34}$$

In what follows, let us suppose that R is a closed real interval. Recall that, according to Remark 2.2, the union of two disjoint fuzzy r_1-partition and r_2-partition is a fuzzy $(r_1 + r_2)$-partition, and this method is very profitable in the construction of new fuzzy r-partitions with higher values of r.

Theorem 4.5. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two disjoint fuzzy r_1-partition and r_2-partition, respectively, and $g \in \mathrm{F}(R, \mathcal{A}_1) \cap \mathrm{F}(R, \mathcal{A}_2)$. Then*

$$V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g), P) \leq \frac{r_1 V_R(\mathcal{F}_{\mathcal{A}_1}(g), P) + r_2 V_R(\mathcal{F}_{\mathcal{A}_2}(g), P)}{r_1 + r_2} \tag{35}$$

for any partition $P \in \mathcal{P}_R$.

Proof. Denote $g^{r_1+r_2} = \mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g)$, $g^{r_1} = \mathcal{F}_{\mathcal{A}_1}(g)$ and $g^{r_2} = \mathcal{F}_{\mathcal{A}_2}(g)$. Analogously, denote the components and basic functions of the F-transform, e.g., $F_i^{r_1+r_2}$ and $A_i^{r_1+r_2}$ denote the i-th component and the basic function of the F-transform, respectively. Now, the equality in (35) can be rewritten as

$$V_R(g^{r_1+r_2}, P) \leq \frac{r_1 V_R(g^{r_1}, P) + r_2 V_R(g^{r_2}, P)}{r_1 + r_2}. \tag{36}$$

Put $n_P = |P|$, $k_1 = |\mathcal{A}_1|$, $k_2 = |\mathcal{A}_2|$ and define $\pi : \{1, \ldots, k_1\} \to \{1, \ldots, k_1 + k_2\}$ and $\rho : \{1, \ldots, k_2\} \to \{1, \ldots, k_1 + k_2\}$ such that $A_s^{r_1} = A_{\pi(s)}^{r_1+r_2}$ and $A_t^{r_2} = A_{\rho(t)}^{r_1+r_2}$

are satisfied for any $s = 1, \dots, k_1$ and $t = 1, \dots, k_2$. Obviously, we have $F_s^{r_1} = F_{\pi(s)}^{r_1+r_2}$ and $F_t^{r_2} = F_{\rho(t)}^{r_1+r_2}$. Now, we have (in the new notation)

$$
\begin{gathered}
|g^{r_1+r_2}(x_{j+1}) - g^{r_1+r_2}(x_j)| = \\
\frac{1}{r_1+r_2}\left|\sum_{i=1}^{k_1+k_2} F_i^{r_1+r_2} A_i^{r_1+r_2}(x_{j+1}) - \sum_{i=1}^{k_1+k_2} F_i^{r_1+r_2} A_i^{r_1+r_2}(x_j)\right| = \\
\frac{1}{r_1+r_2}\left|\left(\sum_{s=1}^{k_1} F_{\pi(s)}^{r_1+r_2} A_{\pi(s)}^{r_1+r_2}(x_{j+1}) + \sum_{t=1}^{k_2} F_{\rho(t)}^{r_1+r_2} A_{\rho(t)}^{r_1+r_2}(x_{j+1})\right)\right. \\
\left. - \left(\sum_{s=1}^{k_1} F_{\pi(s)}^{r_1+r_2} A_{\pi(s)}^{r_1+r_2}(x_j) + \sum_{t=1}^{k_2} F_{\rho(t)}^{r_1+r_2} A_{\rho(t)}^{r_1+r_2}(x_j)\right)\right| = \\
\frac{1}{r_1+r_2}\left|\frac{r_1}{r_1}\left(\sum_{s=1}^{k_1} F_s^{r_1} A_s^{r_1}(x_{j+1}) - \sum_{s=1}^{k_1} F_s^{r_1} A_s^{r_1}(x_j)\right)\right. \\
\left. + \frac{r_2}{r_2}\left(\sum_{t=1}^{k_2} F_t^{r_2} A_t^{r_2}(x_{j+1}) - \sum_{t=1}^{k_2} F_t^{r_2} A_t^{r_2}(x_j)\right)\right| = \\
\frac{1}{r_1+r_2}|r_1(g^{r_1}(x_{j+1}) - g^{r_1}(x_j)) + r_2(g^{r_2}(x_{j+1}) - g^{r_2}(x_j))| \leq \\
\frac{1}{r_1+r_2}(r_1|g^{r_1}(x_{j+1}) - g^{r_1}(x_j)| + r_2|g^{r_2}(x_{j+1}) - g^{r_2}(x_j)|)
\end{gathered}
$$

for any $j = 1, \dots, n_P - 1$. Hence, we obtain

$$
\begin{gathered}
V_R(g^{r_1+r_2}, P) = \sum_{j=1}^{n_P-1} |g^{r_1+r_2}(x_{j+1}) - g^{r_1+r_2}(x_j)| \leq \\
\frac{1}{r_1+r_2}\sum_{j=1}^{n_P-1} (r_1|g^{r_1}(x_{j+1}) - g^{r_1}(x_j)| + r_2|g^{r_2}(x_{j+1}) - g^{r_2}(x_j)|) = \\
\frac{r_1 V_R(g^{r_1}, P) + r_2 V_R(g^{r_2}, P)}{r_1+r_2}.
\end{gathered}
$$

□

Straightforward consequences are contained in the following corollary.

Corollary 4.6. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two disjoint fuzzy r_1-partition and r_2-partition, respectively, and $g \in \mathrm{F}(R, \mathcal{A}_1) \cap \mathrm{F}(R, \mathcal{A}_2)$. Then*

$$
V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g)) \leq \frac{r_1 V_R(\mathcal{F}_{\mathcal{A}_1}(g)) + r_2 V_R(\mathcal{F}_{\mathcal{A}_2}(g))}{r_1+r_2}, \tag{37}
$$

$$
V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g)) \leq \max(V_R(\mathcal{F}_{\mathcal{A}_1}(g)), V_R(\mathcal{F}_{\mathcal{A}_2}(g))). \tag{38}
$$

The second statement of the previous corollary shows that, in general, we cannot ensure a more smoothed function by increasing value of r. More precisely,

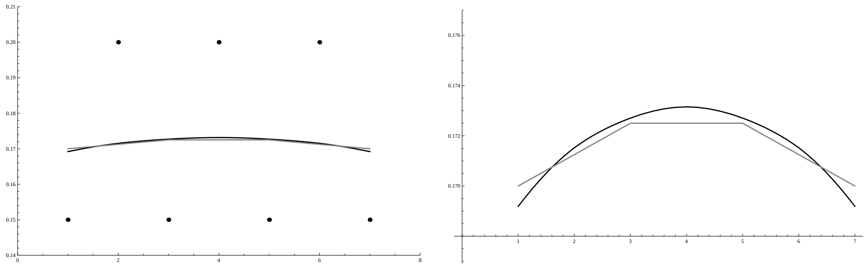

Figure 6: Smoothed functions under the uniform fuzzy 20-partition (black line) and the uniform fuzzy 1-partition (grey line) for $h = 4$.

a higher value of r does not lead generally to a lower value of the total variation. On Fig. 6, one can see a rather artificial example demonstrating that for the same basal fuzzy sets the smoothed function constructed over the uniform fuzzy 1-partition gives a lower value of the total variation than over the uniform fuzzy 20-partition. For a detail, see the right part of Fig. 6.

A consequence of the first statement of the previous corollary is the fact that for a combination of two disjoined fuzzy r-partitions the total variation of the resulted function is more influenced by the fuzzy r-partition with the higher value of r. Hence, the smoothness of the resulted function need not be significantly changed for denser fuzzy r-partitions (consider the black curve in Fig. 6).

Although, the higher smoothness is not generally guaranteed by higher values of r, we can show that, under some assumptions, the total variation of the resulted functions can be higher for denser fuzzy r-partitions. Recall that $V_R(g, P) \le V_R(g, P')$ for any $P, P' \in \mathcal{P}_R$ such that $P \subseteq P'$.

Corollary 4.7. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be disjoint fuzzy r_1-partition and r_2-partition, respectively, and $g \in \mathrm{F}(R, \mathcal{A}_1) \cap \mathrm{F}(R, \mathcal{A}_2)$. If for any $P \in \mathcal{P}_R$ there exist $P', P'' \in \mathcal{P}_R$ such that*

$$V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g), P) \le \min(V_R(\mathcal{F}_{\mathcal{A}_1}(g), P'), V_R(\mathcal{F}_{\mathcal{A}_2}(g), P'')) \qquad (39)$$

then also $V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g)) \le \min(V_R(\mathcal{F}_{\mathcal{A}_1}(g)), V_R(\mathcal{F}_{\mathcal{A}_2}(g)))$.

Proof. As a simple consequence of assumption (39), we obtain

$$V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g), P) \le V_R(\mathcal{F}_{\mathcal{A}_1}(g)) \text{ and } V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g), P) \le V_R(\mathcal{F}_{\mathcal{A}_2}(g))$$

for any $P \in \mathcal{P}_R$. Hence, $V_R(\mathcal{F}_{\mathcal{A}_1 \cup \mathcal{A}_2}(g), P) \le \min(V_R(\mathcal{F}_{\mathcal{A}_1}(g)), V_R(\mathcal{F}_{\mathcal{A}_2}(g)))$ which implies the desired conclusion. □

As one could notice from Fig. 4 and Fig. 5, it is difficult to say precisely which function is smoothed better. Clearly, it depends on further requirements. For example, one can ask for a smoothed function which fits the original data from a sample well, another one is looking for a smoothed function to reduce a white noise from data and so be able to recognize their structure (see Corollary 4.10) and etc. Generally, we can say that, in many cases for the fixed bandwidth h of a basal fuzzy set, a higher value of r (a denser uniform fuzzy r-partition) ensures a more smoothed function, especially, if discrete functions with greater variability are considered.

4.5. Statistical analysis of FT-smoothing filter estimator in the fixed design

In this part, we shall continue our statistical analysis of the F-transform components from the kernel-based nonparametric regression perspective presented in the previous section and apply its results to the statistical analysis of FT-smoothing filter.

Let R be a real interval, $x_1 < \cdots < x_n$ be a sequence of nodes of R and $\mathcal{A}$ be a finite fuzzy r-partition of R determined by $(\mathcal{T}, \mathcal{S})$. Further, let us suppose that $X = \{x_1, \ldots, x_n\}$ is sufficiently dense with respect to $\mathcal{A}$. Finally, put $Y_j = g(x_j) + \varepsilon_j$, $j = 1, \ldots, n$, where g is a (non-random) function and ε_j is a random variable representing the error in x_j with $\mathrm{E}(\varepsilon_j) = 0$, $\mathrm{Var}(\varepsilon_j) = \sigma^2 < \infty$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, and denote

$$\hat{g}_{\mathrm{FT}}(x) = \frac{1}{r} \sum_{i=1}^{k} \Phi_i A_i(x) \tag{40}$$

the FT-smoothing filter estimator of the function g at the point x, where Φ_i are the stochastic F-transform components of the finite random function given by $(x_1, Y_1), \ldots, (x_n, Y_n)$. Since $\mathrm{E}(\Phi_i) = F_i$ for any $i = 1, \ldots, k$, where F_i are the F-transform components of the function g, one can prove easily that

$$\mathrm{E}(\hat{g}_{\mathrm{FT}}(x)) = \mathcal{F}_{\mathcal{A}}(g)(x) \tag{41}$$

for any $x \in R$.

Theorem 4.8. *Let $x \in R$ and $\hat{g}_{\mathrm{NW}}(x)$ denote the NW estimator of a function g at the point x. Then*

$$\mathrm{Bias}(\hat{g}_{\mathrm{FT}}(x)) = \frac{1}{r} \sum_{i=1}^{k} \mathrm{Bias}(\hat{g}_{\mathrm{NW}}(t_i)) A_i(x) + \frac{1}{r} \sum_{i=1}^{k} (g(t_i) - g(x)) A_i(x), \tag{42}$$

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{1}{r^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{Cov}(\Phi_i, \Phi_j) A_i(x) A_j(x). \tag{43}$$

Proof. Let $x \in R$. A simple consequence of $\Phi_i = \hat{g}_{\mathrm{NW}}(t_i)$ (see Proposition 3.2) is

$$\mathrm{E}(\hat{g}_{\mathrm{FT}}(x)) = \frac{1}{r} \sum_{i=1}^{k} \mathrm{E}(\hat{g}_{\mathrm{NW}}(t_i)) A_i(x).$$

Note that $\mathrm{E}(\hat{g}_{\mathrm{NW}}(t_i)A_i(x)) = \mathrm{E}(\hat{g}_{\mathrm{NW}}(t_i))A_i(x)$ follows from the fact that $A_i(x)$ is a constant (i.e. a non-random number) for any $i = 1, \dots, k$. Hence, we obtain

$$\mathrm{Bias}(\hat{g}_{\mathrm{FT}}(x)) = \mathrm{E}(\hat{g}_{\mathrm{FT}}(x)) - g(x) =$$
$$\frac{1}{r}\sum_{i=1}^{k} \mathrm{E}(\hat{g}_{\mathrm{NW}}(t_i) - g(t_i) + g(t_i) - g(x))A_i(x),$$

which gives, after a simple modification, the first statement.

For the variability, we have

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \mathrm{E}(\hat{g}_{\mathrm{FT}}(x)^2) - \left(\mathrm{E}(\hat{g}_{\mathrm{FT}}(x))\right)^2 =$$
$$\frac{1}{r^2}\mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k} \Phi_i\Phi_j A_i(x)A_j(x)\right) - \frac{1}{r^2}\left(\sum_{i=1}^{k}\mathrm{E}(\Phi_i)A_i(x)\right)^2 =$$
$$\frac{1}{r^2}\sum_{i=1}^{k}\sum_{j=1}^{k}(\mathrm{E}(\Phi_i\Phi_j) - \mathrm{E}(\Phi_i)\mathrm{E}(\Phi_j))A_i(x)A_j(x) =$$
$$\frac{1}{r^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(\Phi_i, \Phi_j)A_i(x)A_j(x)$$

and the proof is finished. □

One can see that a defect of the previous theorem is the expression of the covariance of random variables Φ_i and Φ_j in the formula of $\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x))$. In the fixed design, we may find a better description as the following theorem shows. Note that such expression is more or less meaningful for the random design (cf. [21]). An asymptotic expressions of $\mathrm{Cov}(\Phi_i, \Phi_j)$ in the fixed equally spaced design will be presented later. As we have mentioned in Introduction, this expression is an open problem for the random design.

Theorem 4.9. *Let $x \in R$. Then*

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{\sigma^2}{r^2}\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}A_i(x)A_j(x), \tag{44}$$

where

$$a_{ij} = \frac{\sum_{s=1}^{n} A_i(x_s)A_j(x_s)}{\sum_{s=1}^{n} A_i(x_s)\sum_{s=1}^{n} A_j(x_s)}. \tag{45}$$

Proof. According to Theorem 4.8, it is sufficient to show that $\mathrm{Cov}(\Phi_i, \Phi_j) =$

$\sigma^2 a_{ij}$ for any $i, j = 1, \ldots, k$. Thus, for $\mathrm{Cov}(\Phi_i, \Phi_j)$, we have

$$\mathrm{Cov}(\Phi_i, \Phi_j) = \mathrm{E}(\Phi_i, \Phi_j) - \mathrm{E}(\Phi_i)\mathrm{E}(\Phi_j) =$$
$$\mathrm{E}\left(\frac{\sum_{t=1}^n \sum_{s=1}^n Y_t Y_s A_i(x_t) A_j(x_s)}{\sum_{t=1}^n A_i(x_t) \sum_{s=1}^n A_j(x_s)}\right) - \frac{\sum_{t=1}^n \sum_{s=1}^n \mathrm{E}(Y_t)\mathrm{E}(Y_s) A_i(x_t) A_j(x_s)}{\sum_{t=1}^n A_i(x_t) \sum_{s=1}^n A_j(x_s)} =$$
$$\frac{\sum_{t=1}^n \sum_{s=1}^n (\mathrm{E}(Y_t Y_s) - \mathrm{E}(Y_t)\mathrm{E}(Y_s)) A_i(x_t) A_j(x_s)}{\sum_{t=1}^n A_i(x_t) \sum_{s=1}^n A_j(x_s)} =$$
$$\frac{\sum_{t=1}^n \sum_{s=1}^n \mathrm{Cov}(Y_i, Y_j) A_i(x_t) A_j(x_s)}{\sum_{t=1}^n A_i(x_t) \sum_{s=1}^n A_j(x_s)}.$$

According to the assumption on the random variables Y_i and Y_j (see (14) on page 13), we have $\mathrm{Cov}(Y_i, Y_j) = 0$ for any $i \neq j$ (the different random variables are independent) and $\mathrm{Cov}(Y_i, Y_i) = \mathrm{Var}(Y_i) = \sigma^2$. Applying these facts, we obtain

$$\mathrm{Cov}(\Phi_i, \Phi_j) = \sigma^2 \frac{\sum_{s=1}^n A_i(x_s) A_j(x_s)}{\sum_{s=1}^n A_i(x_s) \sum_{s=1}^n A_j(x_s)}. \tag{46}$$

Setting

$$a_{ij} = \frac{\sum_{s=1}^n A_i(x_s) A_j(x_s)}{\sum_{s=1}^n A_i(x_s) \sum_{s=1}^n A_j(x_s)} \tag{47}$$

we obtain $\mathrm{Cov}(\Phi_i, \Phi_j) = \sigma^2 a_{ij}$ for any $i, j = 1, \ldots, k$. □

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

Corollary 4.10. *If $X = \{x_1, \ldots, x_n\}$ be a set of nodes that is 2-dense with respect to $\mathcal{A}$. Then $\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) < \sigma^2$.*

Proof. Put $c = \max\{a_{ij} \mid i, j = 1, \ldots, k\}$. Then

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{\sigma^2}{r^2} \sum_{i=1}^k \sum_{j=1}^k a_{ij} A_i(x) A_j(x) \leq \frac{c\sigma^2}{r^2} \sum_{i=1}^k \sum_{j=1}^k A_i(x) A_j(x) =$$
$$\frac{c\sigma^2}{r^2} \left(\sum_{i=1}^k A_i(x)\right)^2 = \frac{c\sigma^2}{r^2} r^2 = c\sigma^2.$$

Since $c \leq 1$, then $\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) \leq \sigma^2$ holds in general. Let us suppose that X is 2-dense with respect to $\mathcal{A}$. Then we obtain that $c < 1$. In fact, let us suppose that $a_{ij} = 1$ for some $i, j = 1, \ldots, k$. This case arises when

$$\sum_{s=1}^n A_i(x_s) A_j(x_s) = \sum_{s=1}^n A_i(x_s) \sum_{s=1}^n A_j(x_s)$$

or equivalently

$$\sum_{\substack{s,t=1\\ s\neq t}}^{n} A_i(x_s)A_j(x_t) = 0.$$

But this is impossible with regard to the assumption on the 2-density of X. Hence, $a_{ij} < 1$ for any $i, j = 1, \dots, k$. Note that the assumption on X to be 2-dense is important mainly for the case $i = j$. □

Now, we shall study the asymptotic properties of the FT-smoothing filter estimator. In order to use the results about the asymptotic behavior of the NW estimator and to compare both smoothers, we shall express the FT-smoothing filter estimator in terms of kernels. 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 (see Proposition 2.3), then

$$\hat{g}_{\mathrm{FT}}(x) = \frac{1}{r}\sum_{i=1}^{k}\Phi_i S(x - t_i) = \frac{1}{rK(0)}\sum_{i=1}^{k}\Phi_i K(x - t_i). \tag{48}$$

Since $\sum_{j=1}^{n} K(x - t_j) = rK(0)$ for any $x \in R$ and $\Phi_i = \hat{g}_{\mathrm{NW}}(t_i)$, we may introduce a uniform kernel r'-partition determined by $(\mathcal{T}, K)$ with $r' = rK(0)$ and the FT-smoothing filter estimator defined in terms of kernels by

$$\hat{g}_{\mathrm{FT}}(x) = \frac{1}{r'}\sum_{i=1}^{k}\Phi_i K(x - t_i). \tag{49}$$

Let us stress that the interpretation of values of r' is rather different from the values of r for the fuzzy r-partitions. While the value of r roughly speaking refers to a number of active basic functions (especially, basal fuzzy sets) for an element $x \in R$ over which the evaluation of the FT-smoothing filter estimator is made (cf. [30]), the value of r' is only an abstract value strongly depending on the high $K(0)$ of the kernel K. Obviously, r' becomes interpretable for $r'/K(0)$. For simplicity, we shall omit the prime in r' and write only r in the following text.

Before commencing our study of asymptotic properties of the FT-smoothing filter estimator in the fixed equally spaces design, let us recall that $Y_j = g(x_j) + \varepsilon_j$, $j = 1, \dots, n$, where $x_j = a + j(b - a)/n$, g is a (non-random) function and ε_j is a random variable representing the error in x_j with $\mathrm{E}(\varepsilon_j) = 0$, $\mathrm{Var}(\varepsilon_j) = \sigma^2 < \infty$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. Further, let K be a symmetric kernel with $K(x) = 0$ for any $x \notin [-1, 1]$ and $\mathcal{A}_n$ denote a uniform kernel r_n-partition of a real interval R determined by $(\mathcal{T}_{k_n}, K_{h_n})$, where k_n is the number of nodes over which the kernel r_n-partition is constructed and h_n is the bandwidth of K_{h_n}. Finally, define

$$\alpha(x, y) = \int_{-1}^{1} K(z)K\left(\frac{x - y}{h} + z\right) dz \tag{50}$$

and put $u_n = u = t_{i+1} - t_i$ for $t_i, t_{i+1} \in \mathcal{T}_{k_n}$. Obviously, $\alpha(x, x) = R(K)$. Notice that $u_n \leq h_n$, otherwise, $\mathcal{A}_n$ cannot be a kernel r_n-partition since $\sum_{i=1}^{k_n} A_i(x) = 0$ for some $x \in [a, b]$.

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

(FA1) The functions g 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 $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

(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) $\mathcal{A}_n$ is a uniform kernel r_n-partition of $[a + h_n, b - h_n]$ for all $n \geq n_0$, where n_0 is fixed.

(FA6) $k = k_n$, $r = r_n$ and $u = u_n$ are sequences satisfying $r_n \to \infty$, $nh_n^3 k_n^2 \to \infty$, $h_n^2 r_n^{-1} u_n^{-1} \to 0$ and $k_n^2 r_n^{-2} n^{-1} \to 0$ as $n \to \infty$.

(FA7) $\alpha(x, y)$ has continuous partial derivatives in $[a, b]$.

The condition (FA6) states the important convergences under which the asymptotic expressions of the bias and variance of $\hat{g}_{\mathrm{FT}}(x)$ can be derived. The first convergence $r \to \infty$ is a natural consequence of $h \to 0$. Moreover, if $h \to 0$, then the number of basic functions $k \to \infty$. The convergence $nh^3k^2 \to \infty$ is an analogy of $nh \to \infty$ and seems to be acceptable. As we have mentioned $u \leq h$. The rate hu^{-1} approximately characterizes the number $|d_{\mathcal{A}}(x)|$ of active basic functions. The convergence $h^2r^{-1}u^{-1} \to 0$ is an important assumption ensuring that $\hat{g}_{\mathrm{FT}}(x)$ is an asymptotically unbiased estimator (see also Remark 2.3). Notice that a higher number k of basic functions in the closed interval implies a greater sum in the r-Ruspini condition. The last convergence $k^2r^{-2}n^{-1} \to 0$ says that the number of data has to be significantly greater than the rate k^2r^{-2} and this can be insured.

Note that a problem can arise if one wants to verification of (FA7) for some type of kernel. Unfortunately, this assumption is important to simplify the expression of the coefficients a_{ij} in the proof of $\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x))$. We checked up (FA7) using numerical methods and this assumption seems to be satisfied for the triangle and raised cosine kernels. A mathematical verification is still an open problem.

An asymptotic characterization of Bias and Var of the estimator $\hat{g}_{\mathrm{FT}}(x)$ can be described as follows.

Theorem 4.11. *Let the assumptions (FA1)-(FA7) be satisfied and* $\frac{1}{n} \in O\left(\frac{1}{r}\right)$*, then*

$$\operatorname{Bias}(\hat{g}_{\mathrm{FT}}(x)) = \frac{h^2\mu_2(K)}{ru} g''(x) + o\left(\frac{h^2}{ru} + h^2\right) + O\left(\frac{1}{r}\right), \tag{51}$$

$$\operatorname{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{k^2\sigma^2}{r^2nh(b-a)} R(K) + o\left(\frac{k^2}{nr^2} + \frac{1}{nh^3r^2}\right), \tag{52}$$

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

Proof. According to Theorem 3.3 and Theorem 4.8, we have

$$\begin{aligned}
\operatorname{Bias}(\hat{g}_{\mathrm{FT}}(x)) =& \frac{1}{r}\sum_{i=1}^{k} \operatorname{Bias}(\hat{g}_{\mathrm{NW}}(t_i))A_i(x) + \frac{1}{r}\sum_{i=1}^{k}(g(t_i) - g(x))A_i(x) = \\
& \frac{h^2\mu_2(K)}{2r}\sum_{i=1}^{k} g''(t_i)A_i(x) + \frac{1}{r}\sum_{i=1}^{k}(g(t_i) - g(x))A_i(x) \\
& + O\left(\frac{1}{n}\right) + o(h^2).
\end{aligned} \tag{53}$$

where $\frac{1}{r}\sum_{i=1}^{k} A_i(x) = 1$ is used. First, let us consider

$$\frac{1}{r}\sum_{i=1}^{k} g''(t_i)A_i(x) = \frac{u}{hur}\sum_{i=1}^{k} g''(t_i)K\left(\frac{t_i - x}{h}\right),$$

where $u = t_{i+1} - t_i$. Note that $K\left(\frac{t_i-x}{h}\right) = K\left(\frac{x-t_i}{h}\right)$ follows from the symmetry of K. Put $\mathcal{A} = \mathcal{A}_n$. Since $h \to 0$, then $u \to 0$ and we obtain

$$\frac{u}{h}\sum_{i=1}^{k} g''(t_i)K\left(\frac{t_i - x}{h}\right) = \frac{1}{h}\int_{t_1}^{t_k} g''(y)K\left(\frac{y-x}{h}\right)dy + O(u). \tag{54}$$

In fact, put $f(y) = g''(y)K\left(\frac{y-x}{h}\right)$. Obviously, f is a continuous function and $f(t_i) = 0$ for any $t_i \notin d_{\mathcal{A}}(x)$ and $|d_{\mathcal{A}}(x)|u \le 2h$.[13] A straightforward consequence of (FA5) and (FA6) is the existence of $t_{i_0} \in \mathcal{T}_k$ such that $t_i < t_{i_0}$ holds for any $i \in d_{\mathcal{A}}(x)$. Applying this fact and the mean value theorem for integration we obtain

$$\begin{aligned}
\left|\frac{1}{h}\sum_{i=1}^{k} f(t_i)u - \frac{1}{h}\int_{t_1}^{t_k} f(y)dy\right| &= \left|\frac{1}{h}\sum_{i\in d_{\mathcal{A}}(x)} f(t_i)u - \frac{1}{h}\sum_{i\in d_{\mathcal{A}}(x)}\int_{t_i}^{t_{i+1}} f(y)dy\right| = \\
\frac{1}{h}\left|\sum_{i\in d_{\mathcal{A}}(x)} f(t_i)u - \sum_{i\in d_{\mathcal{A}}(x)} f(\xi_i)u\right| &\le \frac{1}{h}\sum_{i\in d_{\mathcal{A}}(x)} |f(t_i) - f(\xi_i)|u \le \\
\frac{1}{h}\sum_{i\in d_{\mathcal{A}}(x)} (cu)u &\le \frac{c}{h}(|d_{\mathcal{A}}(x)|u)u \le \frac{c}{h}2hu = 2cu
\end{aligned}$$

[13] Recall that $t_i \in d_{\mathcal{A}}(x)$ if and only if $t_i \in]x-h, x+h[$.

where $\xi_i \in [t_i, t_{i+1}]$ and c is a constant which existence follows from the continuity of f on $[a,b]$. Now, we may set up the considered integral as follows

$$\frac{1}{h}\int_{t_1}^{t_k} g''(y)K\left(\frac{y-x}{h}\right)dy + O(u) =$$
$$\int_{(t_1-x)/h}^{(t_k-x)/h} g''(x+hz)K(z)dz + O(u) = \int_{-1}^{1} g''(x+hz)K(z)dz + O(u),$$

where we put $z = (y-x)/h$ and use $(b-x)/h \geq 1$, $(a-x)/h \leq -1$ and $K(x) = 0$ for $x \notin (-1,1)$. Using (FA1), we may use the Taylor expansion of $g''(x+hz)$ as follows

$$g''(x+hz) = g''(x) + g^{(3)}(x)hz + \frac{g^{(4)}(x)}{2}(hz)^2 + o(h^2).$$

Hence, we have

$$\int_{-1}^{1} g''(x+hz)K(z)dz + O(u) = g''(x)\int_{-1}^{1} K(z)dz +$$
$$g^{(3)}(x)h\int_{-1}^{1} zK(z)dz + g^{(4)}(x)h^2\int_{-1}^{1} z^2K(z)dz + o(h^2) + O(u) =$$
$$g''(x) + \frac{h^2\mu_2(K)}{2}g^{(4)}(x) + o(h^2) + O(u)$$

and thus

$$\begin{aligned}\frac{1}{r}\sum_{i=1}^{k} g''(t_i)A_i(x) = \frac{1}{ur}\frac{u}{h}\sum_{i=1}^{k} g''(t_i)K\left(\frac{t_i-x}{h}\right) =\\ \frac{1}{ur}\left(g''(x) + \frac{h^2\mu_2(K)}{2}g^{(4)}(x) + o(h^2) + O(u)\right) =\\ \frac{1}{ur}\left(g''(x) + \frac{h^2\mu_2(K)}{2}g^{(4)}(x)\right) + o\left(\frac{h^2}{ur}\right) + O\left(\frac{1}{r}\right).\end{aligned} \tag{55}$$

Further, let us consider

$$\frac{1}{r}\sum_{i=1}^{k}(g(t_i) - g(x))A_i(x) = \frac{u}{rhu}\sum_{i=1}^{k}(g(t_i) - g(x))K\left(\frac{t_i-x}{h}\right).$$

Analogously, one may derive

$$\frac{1}{r}\sum_{i=1}^{k}(g(t_i) - g(x))A_i(x) = \frac{h^2\mu_2(K)}{2ru}g''(x) + o\left(\frac{h^2}{ru}\right) + O\left(\frac{1}{r}\right). \tag{56}$$

Plugging (55) and (56) to the formula (53), we obtain

$$\begin{aligned}
&\mathrm{Bias}(\hat{g}_{\mathrm{FT}}(x)) = \\
&\quad \frac{h^2\mu_2(K)}{2}\left(\frac{1}{ru}\left(g''(x)+\frac{h^2\mu_2(K)}{2}g^{(4)}(x)\right)+o\left(\frac{h^2}{ur}\right)+O\left(\frac{1}{r}\right)\right)+ \\
&\quad \frac{h^2\mu_2(K)}{2ru}g''(x)+o\left(\frac{h^2}{ru}\right)+O\left(\frac{1}{r}\right)+O\left(\frac{1}{n}\right)+o(h^2)= \\
&\quad \frac{h^2\mu_2(K)}{ru}g''(x)+o\left(\frac{h^2}{ru}+h^2\right)+O\left(\frac{1}{r}\right),
\end{aligned}$$

where $\frac{1}{n}\in O\left(\frac{1}{r}\right)$ is applied and $\frac{h^4\mu_2(K)^2}{4ur}g^{(4)}(x)\in o\left(\frac{h^2}{ru}\right)$ holds true, since $g^{(4)}(x)$ is bounded according to (FA1).

In order to find the asymptotic expression of the variance of $\hat{g}_{\mathrm{FT}}(x)$, let us start with the asymptotic expression of coefficients a_{ij} defined in Theorem 4.9. Obviously, we have

$$a_{ij}=\frac{\left(\frac{b-a}{n}\right)^2\frac{1}{h^2}\sum_{s=1}^n K\left(\frac{x_s-t_i}{h}\right)K\left(\frac{x_s-t_j}{h}\right)}{\left(\frac{b-a}{n}\right)^2\frac{1}{h^2}\sum_{s=1}^n K\left(\frac{x_s-t_i}{h}\right)\sum_{s=1}^n K\left(\frac{x_s-t_j}{h}\right)}. \tag{57}$$

Put

$$T_i=\left(\frac{b-a}{hn}\right)\sum_{s=1}^n K\left(\frac{x_s-t_i}{h}\right), \tag{58}$$

$$T_{ij}=\left(\frac{b-a}{hn}\right)\sum_{s=1}^n K\left(\frac{x_s-t_i}{h}\right)K\left(\frac{x_s-t_j}{h}\right). \tag{59}$$

To simplify T_i, define $f(y)=\frac{1}{h}K(\frac{y-t_i}{h})$ and put $x_0=a$. Then

$$\begin{aligned}
&\left|T_i-\int_a^b f(y)dy\right| = \left|\frac{b-a}{n}\sum_{s=1}^n f(x_s)-\int_a^b f(y)dy\right| \le \\
&\sum_{s=1}^n\left|f(x_s)\left(\frac{b-a}{n}\right)-\int_{x_{s-1}}^{x_s} f(y)dy\right| = \sum_{s=1}^n\left|f(x_s)\left(\frac{b-a}{n}\right)-f(\xi_s)\left(\frac{b-a}{n}\right)\right| = \\
&\sum_{s=1}^n|f(x_s)-f(\xi_s)|\left(\frac{b-a}{n}\right)\le\sum_{s=1}^n c\left(\frac{b-a}{n}\right)^2\le\frac{c(b-a)^2}{n}
\end{aligned}$$

where $\xi_s\in[x_{s-1},x_s]$ is the mean value[14] and $|f(x_s)-f(\xi_s)|\le c|x_s-\xi_s|\le c\frac{b-a}{n}$ follows from the continuity of f in $[a,b]$. Hence, we obtain

$$T_i=\frac{1}{h}\int_a^b K\left(\frac{y-t_i}{h}\right)dy+O\left(\frac{1}{n}\right). \tag{60}$$

[14] Apply the first mean value theorem for integration.

This integral can be rewritten as

$$\frac{1}{h}\int_a^b K\left(\frac{y-t_i}{h}\right)dy = \int_{\frac{a-t_i}{h}}^{\frac{b-t_i}{h}} K(z)dz = \int_{-1}^{1} K(z)dz = 1,$$

where $z = (y-t_i)/h$, $(b-t_i)/h \geq 1$ and $(a-t_i)/h \leq -1$ and $K(x) = 0$ for any $z \notin (-1,1)$ are applied. Thus, we obtain $T_i = 1 + O(n^{-1})$. Analogously, after a bit technical manipulation with terms inside T_{ij}, one can derive

$$T_{ij} = \frac{1}{h}\int_a^b K\left(\frac{y-t_i}{h}\right)K\left(\frac{y-t_j}{h}\right)dy + O\left(\frac{1}{n}\right).$$

The expression of T_{ij} can be rewritten as

$$\int_{(a-t_i)/h}^{(b-t_i)/h} K(z)K\left(\frac{t_i-t_j}{h}+z\right)dz + O\left(\frac{1}{n}\right) = \int_{-1}^{1} K(z)K\left(\frac{t_i-t_j}{h}+z\right)dz + O\left(\frac{1}{n}\right),$$

where $z = (y-t_i)/h$, $(b-t_i)/h \geq 1$ and $(a-t_i)/h \leq -1$ are applied. Thus, $T_{ij} = \alpha(t_i,t_j) + O(n^{-1})$, where $\alpha(x,y)$ is defined in (50). Hence, we have

$$a_{ij} = \frac{(b-a)}{nh}\frac{T_{ij}}{T_iT_j} = \frac{(b-a)\left(\alpha(t_i,t_j)+O\left(\frac{1}{n}\right)\right)}{nh\left(1+O\left(\frac{1}{n}\right)\right)^2} = \frac{(b-a)\left(\alpha(t_i,t_j)+O\left(\frac{1}{n}\right)\right)}{nh\left(1+O\left(\frac{1}{n}\right)\right)} = \frac{(b-a)}{nh}\left(\alpha(t_i,t_j)+O\left(\frac{1}{n}\right)\right) = \frac{(b-a)}{nh}\alpha(t_i,t_j) + o\left(\frac{1}{nh}\right).$$

Plugging the results to (44), we obtain

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{\sigma^2}{r^2h^2}\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_j-x}{h}\right) = \frac{\sigma^2}{r^2h^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\frac{(b-a)}{nh}\alpha(t_i,t_j)+o\left(\frac{1}{nh}\right)\right)K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_j-x}{h}\right) = \frac{\sigma^2}{r^2h^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(b-a)}{nh}\alpha(t_i,t_j)K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_j-x}{h}\right) + o\left(\frac{1}{nh^3r^2}\right) = \frac{k^2\sigma^2\left(\frac{b-a}{kh}\right)^2}{nr^2h(b-a)}\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha(t_i,t_j)K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_j-x}{h}\right) + o\left(\frac{1}{nh^3r^2}\right).$$

Then

$$\left(\frac{b-a}{hk}\right)^2 \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha(t_i,t_j) K\left(\frac{t_i-x}{h}\right) K\left(\frac{t_j-x}{h}\right) =$$

$$\frac{1}{h^2}\int_a^b\int_a^b \alpha(u,v)K\left(\frac{u-x}{h}\right)K\left(\frac{v-x}{h}\right)dudv + O\left(\frac{1}{k^2}\right) =$$

$$\int_{(a-u)/h}^{(b-u)/h}\int_{(a-v)/h}^{(b-v)/h} \alpha(x+yh, x+y'h)K(y)K(y')dydy' + O\left(\frac{1}{k^2}\right) =$$

$$\int_{-1}^{1}\int_{-1}^{1} \alpha(x+yh, x+y'h)K(y)K(y')dydy' + O\left(\frac{1}{k^2}\right).$$

Note that $|f(t_i,t_j) - f(\xi_i,\xi_j)| \le c|t_i-\xi_i||t_j-\xi_j| \le c(b-a)^2/k^2$ is applied to obtain $O(k^{-2})$, where $f(t_i,t_j) = \alpha(t_i,t_j)K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_j-x}{h}\right)$ is a continuous function (see (FA7)). According to (FA7), we may use the Taylor expansion

$$\alpha(u,v) = \alpha(x+yh, x+y'h) = \alpha(x,x) + \alpha_u(x,x)yh + \alpha_v(x,x)y'h + o(h).$$

Then

$$\int_{-1}^{1}\int_{-1}^{1} \alpha(x+uh, x+vh)K(y)K(y')dydy' + O\left(\frac{1}{k^2}\right) =$$

$$\int_{-1}^{1}\int_{-1}^{1} (\alpha(x,x) + \alpha_u(x,x)yh + \alpha_v(x,x)y'h + o(h))K(y)K(y')dydy' + O\left(\frac{1}{k^2}\right) =$$

$$\alpha(x,x) + \alpha_u(x,x)h\int_{-1}^{1}\int_{-1}^{1} yK(y)K(y')dydy' +$$

$$\alpha_v(x,x)h\int_{-1}^{1}\int_{-1}^{1} y'K(y)K(y')dydy' + o(h) + O\left(\frac{1}{k^2}\right) =$$

$$\alpha(x,x) + o(h) + O\left(\frac{1}{k^2}\right).$$

Plugging this result to the expression of $\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x))$, we obtain

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{k^2\sigma^2}{nhr^2(b-a)}\left(\alpha(x,x) + o(h) + O\left(\frac{1}{k^2}\right)\right) + o\left(\frac{1}{nh^3r^2}\right) =$$

$$\frac{k^2\sigma^2}{nhr^2(b-a)}R(K) + o\left(\frac{k^2}{nr^2} + \frac{1}{nh^3r^2}\right),$$

where we use $n^{-1}h^{-1}r^{-2} \in o(n^{-1}h^{-3}r^{-2})$. □

Note that if one assumes the unit interval as $[a,b]$, then formula (52) can be rewritten as

$$\mathrm{Var}(\hat{g}_{\mathrm{FT}}(x)) = \frac{k^2\sigma^2}{r^2nh}R(K) + o\left(\frac{k^2}{nr^2} + \frac{1}{nh^3r^2}\right). \tag{61}$$

The assumptions on $k \to \infty$, the bandwidth $h \to 0$ and $h^2 r^{-1} u^{-1} \to 0$ imply that $\hat{g}_{\mathrm{FT}}(x)$ is an asymptotically unbiased estimator of g similarly to the NW estimator. Let us suppose that $r_n u_n / K_{h_n}(0) = h_n$ (clearly, the bandwidth h_n is the same for a basal fuzzy set and its kernel counterpart). Since $K_h(0) = K(0)/h$, then omitting the indexes for simplicity, one may derive $ru/K(0) = 1$ and thus $ru = K(0)$. This allows us to rewrite the bias of $\hat{g}_{\mathrm{FT}}(x)$ as

$$\mathrm{Bias}(\hat{g}_{\mathrm{FT}}(x)) = \frac{h^2 \mu_2(K)}{K(0)} g''(x) + o\left(\frac{h^2}{K(0)} + h^2\right) + O\left(\frac{1}{r}\right).$$

In comparison with the bias of $\hat{g}_{\mathrm{NW}}(x)$ obtained by the NW estimator, one can see that the rate of the convergence in the case of the FT-smoothing filter estimator is slower than in the case of the NW estimator, since $h^2/K(0)+h^2 \notin o(h^2)$ 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 estimator may be significantly smaller than for the NW estimator, if a larger 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

$$\mathrm{AMSE}(\hat{g}_{\mathrm{FT}}(x)(x)) = \frac{k^2 \sigma^2}{r^2 n h (b-a)} R(K) + \frac{h^4 \mu_2(K)^2}{r^2 u^2} g''(x)^2. \tag{62}$$

The optimal value of bandwidth h_{AMSE} can be derived putting to zero the derivative of AMSE with respect to h. By a simple calculation we obtain

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

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 $i = 1, \ldots, k$, we can simply deduce $ku \geq b - a$. One can notice that $ku \approx b - a$ for small h and

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

Comparing this result with that provided by the NW estimator (see (22) on page 15), we obtain an approximated equality

$$h_{\mathrm{AMSE}}^{\mathrm{FT}} \approx 0.76 h_{\mathrm{AMSE}}^{\mathrm{NW}}, \tag{63}$$

where $0,76 \approx \sqrt[5]{1/4}$. Thus, a lower asymptotic bandwidth for the FT-smoothing filter is needed to obtained an optimal model of an unknown function which would correspond to the NW model.

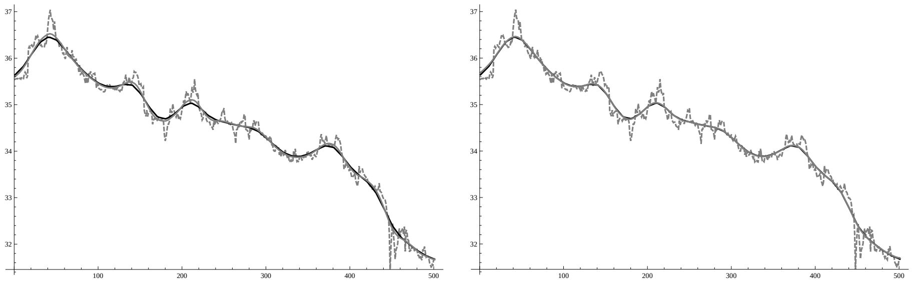

Figure 7: Comparison of $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left) and $h^{\mathrm{NW}} = 1/0.76h^{\mathrm{FT}}$ (right) for $h^{\mathrm{FT}} = 20$ and $r = 2$.

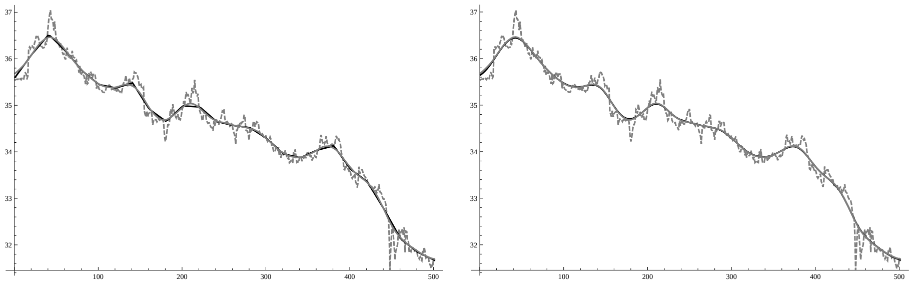

Figure 8: Comparison of the smoothness of resulted functions for $h^{\mathrm{NW}} = 1/0.76h^{\mathrm{FT}}$ with $h^{\mathrm{FT}} = 20$ and $r = 1$ (left) and $r = 8$ (right)

5. Practical comparison of the FT-smoothing filter and NW estimators

For a simple comparison, we chose data sets consisting of 500 and 75 daily quotations of CZK/EUR exchange rate randomly selected from the original time series covering the last eight years.

On Fig. 7, one can see a comparison of the resulted functions obtained by the NW estimator (grey line) and by the FT-smoothing filter estimator (black line), when we use $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left) and $1/0.76h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (right) with $h^{\mathrm{FT}} = 20$ 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. 8, we demonstrate the smoothing property which is clearly dependent on the level of r for the fixed values $h^{\mathrm{FT}} = 20$ and $h^{\mathrm{NW}} = 20/0.76$. The both approaches give us 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 the 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

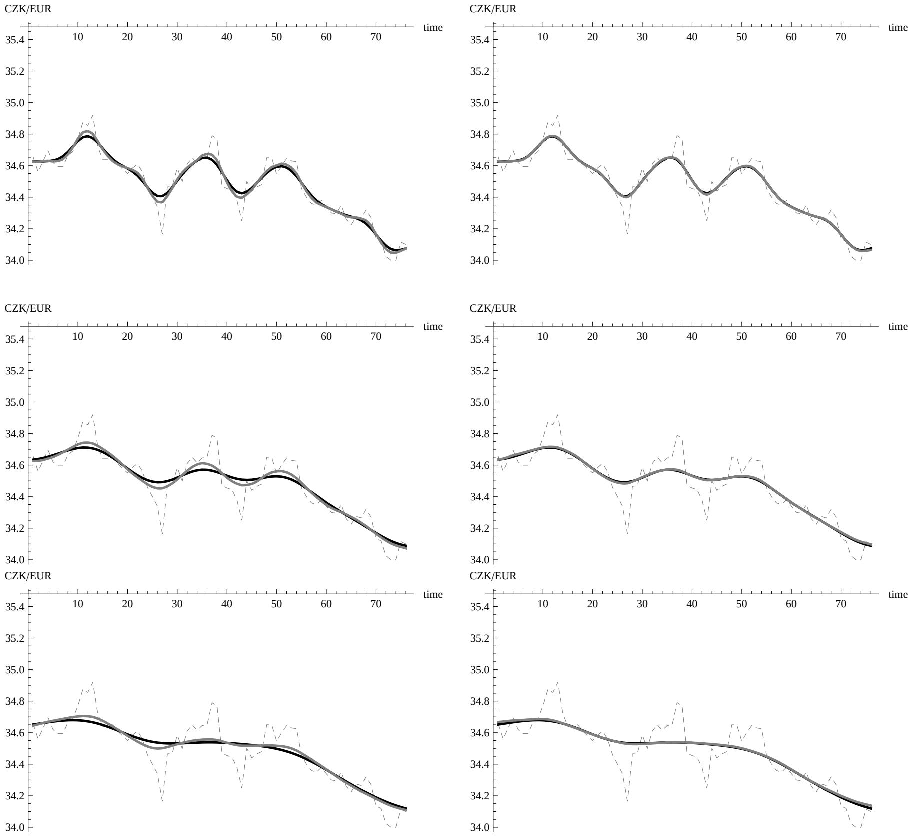

Figure 9: Smoothed functions for financial data consisting of 75 daily quotations of CZK/EUR exchange rate - $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left column) and $h^{\mathrm{NW}} = 1/0.76h^{\mathrm{FT}}$ (right column), $h^{\mathrm{FT}} = 4, 8, 12$.

the function values using NW estimator. On Fig. 9, we present in the first column further smoothed functions by the NW estimator (grey line) and the FT-smoothing filter (black line) assuming the identical bandwidth $h^{\mathrm{FT}} = 4, 8, 12$ with $r = h^{\mathrm{FT}}$. The results assuming the correction on the bandwidth for the NW estimator introduced in (63) are then presented in the second column. Again both approaches provide similar behavior with respect to the same bandwidth (the FT-smoothing filter estimator gives more smoothed functions) and become nearly identical after the proposed correction. Moreover, the number of active values for the kernel is still greater than the number of basic functions used for the evaluation of the FT-smoothing filter estimator.

Summarizing our observation, the FT-smoothing filter can be advantageously used in cases when larger numbers of data are considered to reduce the model complexity of the 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 [12] and is based on the F-transform. To obtain better results of the smoothing procedure, analogously as in [30], 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 fuzzy partitions do not satisfy the Ruspini condition in general. Over such fuzzy r-partitions, the components of the direct discrete F-transform and the FT-smoothing filter were introduced including their stochastic versions. The FT-smoothing filter was defined as a combination of the direct discrete F-transform introduced in [24] and a formula for the inverse continuous F-transform proposed in [30]. An interesting relation between the components of the stochastic F-transform and the Nadaraya-Watson (NW) estimator were proved. Approximation, smoothing and statistical properties of this filter such as the reduction of the white noise, Bias, Var and AMSE were studied. Theoretical results were commented in some details and demonstrated by figures including a comparison with the results obtained by the NW estimator.

One could observe that the FT-smoothing filter (estimator) performs similarly to other filters, mainly, to the NW estimator that is based on kernels and belongs among the traditional smoothing methods of the kernel regression. For example, smaller values of bandwidth h leads to a better approximation which is one of the natural properties of kernel based filters. Further, the higher values of r tend to increase the smoothness of the resulted function. The combination of both parameters allows a user to extend the control over the smoothing procedure. This is a difference against the original definition of the FT-smoothing filter in [12] and also for the NW estimator. For sample data, a reduction of the white noise is ensured assuming a denser sets of data with respect to the fuzzy r-partition. Practically, all samples satisfied this condition and thus a reduction of the white noise is automatically afforded by the FT-smoothing filter. Finally, a valuable relation $0.76h^{\mathrm{NW}} = h^{\mathrm{FT}}$ between the bandwidth h^{NW} of the NW estimator and the bandwidth h^{FT} of the FT-smoothing filter estimator allows us to use estimates of the optimal value of h^{NW} described in the literature related to the kernel smoothing to derive an optimal value of h^{FT}.

Summarizing all mentioned properties the FT-smoothing filter (estimator) clearly belongs to the category of filters based on kernels and plays an analogous role as the finite mixture models which use the mixture distributions to represent the probability distribution of observations in the overall population. An advantage of the FT-smoothing filter is its simpler model complexity against the model complexity of the smoother based on kernels. We can conclude that the FT-smoothing filter provides another useful application of the F-transform which states besides the smoother based on the kernels.

In the future we would like to continue in the investigation of the properties of the FT-smoothing filter defined over F-transform components derived from polynomial. It would be interesting to compare the new results with those which are well known for the local polynomial kernel estimators.

Acknowledgement

The authors wish to thank the anonymous reviewers for their helpful comments and suggestions that have improved the paper.

References

[1] A. Assenza, M. Valle, and M. Verleysen. A comparative study of various probabilty density estimation methods for data analysis. *Int. J. of Computational Intelligence Systems*, 1(2):188–201, 2009.

[2] M. Daňková and R. Valášek. Full fuzzy transform and the problem of image fusion. *J. Electr. Eng.*, 57(7):82–84, 2006.

[3] M. Daňková and M. Štěpnička. Fuzzy transform as an additive normal form. *Fuzzy Sets Syst.*, 157(8):1024–1035, 2006.

[4] F. Di Martino, V. Loia, I. Perfilieva, and S. Sessa. An image coding/decoding method based on direct and inverse fuzzy transforms. *Int. J. Approx. Reasoning*, 48(1):110–131, 2008.

[5] F. di Martino, V. Loia, and S. Sessa. Fuzzy transforms method and attribute dependency in data analysis. *Inf. Sci.*, 180(4):493–505, 2010.

[6] F. Di Martino, V. Loia, and S. Sessa. A segmentation method for images compressed by fuzzy transforms. *Fuzzy Sets Syst.*, 161(1):56–74, 2010.

[7] D. Dubois and H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9:613–626, 1978.

[8] R.L. Eubank. *Spline smoothing and nonparametric regression*. Statistics, textbooks and monographs. M. Dekker, 1988.

[9] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman&Hall/CRC Monographs on Statistics & Applied Probability, 1996.

[10] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics. Springer-Verlag, 2005.

[11] W. Hölder, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and semiparametric models*. Springer-Verlag, Berlin Heidelberg, 2004.

[12] M. Holčapek and T. Tichý. A probability density function estimation using F-transform. *Kybernetika*, 46(3):447– 458, 2010.

[13] G.J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, New Jersey, 1995.

[14] I. Koo and R.M. Kil. Model selection for regression with continuous kernel functions using the modulus of continuity. *Journal of Machine Learning Research*, 9:2607–2633, 2008.

[15] E.J. Kostelich and J.A. Yorke. Noise reduction: Finding the simplest dynamical system consistent with the data. *Physica*, D 41:183–196, 1990.

[16] Hui-Hsiung Kuo. *White noise distribution theory*. CRC Press, 1996.

[17] Ch.S. Lee, Y.H. Kuoa, and P.T. Yub. Weighted fuzzy mean filters for image processing. *Fuzzy Sets Syst.*, 89(2):157–180, 1997.

[18] G. J. McLachlan and D. Peel. *Finite mixture models*. John Wiley & Sons, Inc., 2000.

[19] M. Nachtegael, D. Van der Weken, D. Van De Ville, and E.E. Kerre, editors. *Fuzzy Filters for Image Processing*. Studies in Fuzziness and Soft Computing. Springer-Verlag, 2003.

[20] E.A. Nadaraya. On estimating regression. *Theory Probab. Appl.*, 9(1):141–142, 1964.

[21] A. Pagan and A. Ullah. *Nonparametric Econometrics*. Cambridge University Press, New York, 1999.

[22] I. Perfilieva. Fuzzy transforms. Peters, James F. (ed.) et al., Transactions on Rough Sets II. Rough sets and fuzzy sets. Berlin: Springer. Lecture Notes in Computer Science 3135. Journal Subline, 63-81 (2004), 2004.

[23] I. Perfilieva. Fuzzy transforms and their applications to image compression. Bloch, Isabelle (ed.) et al., Fuzzy logic and applications. 6th international workshop, WILF 2005, Crema, Italy, September 15–17, 2005. Revised selected papers. Berlin: Springer. Lecture Notes in Computer Science 3849. Lecture Notes in Artificial Intelligence, 19-31 (2006), 2006.

[24] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets syst.*, 157(8):993–1023, 2006.

[25] I. Perfilieva and M. Daňková. Towards f-transform of a higher degree. In *Proceedings of IFSA/EUSFLAT 2009*, pages 585–588, Lisbon, Portugal, 2009.

[26] I. Perfilieva, V. Novák, and A. Dvořák. Fuzzy transform in the analysis of data. *Int. J. Approx. Reasoning*, 48(1):36–46, 2008.

[27] E.H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[28] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, Lodon, 1986.

[29] J. S. Simonoff. *Smoothing methods in statistics*. Springer-Verlag, New York, 1996.

[30] L. Stefanini. Fuzzy transform and smooth function. In *Proceedings of IFSA/EUSFLAT 2009*, pages 579–584, Lisbon, Portugal, 2009.

[31] 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.

[32] M. Štěpnička. Fuzzy transformation and its applications in a/d converter. *J. Electr. Eng.*, 54(12):72–75, 2003.

[33] M. Štěpnička and R. Valášek. Fuzzy transforms and their application to wave equation. *J. Electr. Eng.*, 55(12):7–10, 2004.

[34] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman&Hall/CRC Monographs on Statistics & Applied Probability, London, 1995.

[35] G.S. Watson. Smooth regression analysis. *Sankhya*, Series A, 26:359–372, 1964.

[36] J. Yeh. *Lectures on real analysis.* Singapore: World Scientific. xvi, 548 p., 2000.