2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Discrete Fuzzy Transform of Higher Degree

Michal Holčapek and Tomáš Tichý

***Abstract*— In this paper, we reformulate the fuzzy transform of higher degree (F^m-transform) proposed originally for an approximation of continuous functions to the discrete case. We introduce two types of F^m-transform which components are defined using polynomials in the first case and using specific values of these polynomials in the second case. We provide an analysis of basic properties of F^m-transform.**

I. Introduction

FFUZZY TRANSFORM is a special soft computing technique proposed by Perfilieva in [1] (see also [2]) that has many applications in various fields, for example in data analysis, image processing, approximate solution of differential equations, time series analysis, non-parametric regression, and elsewhere (for details, we refer to [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]). Recall that the core of fuzzy transform (F-transform) consists in partitioning of a real interval using fuzzy sets.

The F-transform has two phases: direct and inverse. The *direct* F-transform transforms a bounded real function f to a finite vector of real numbers (components of F-transform). The *inverse* F-transform sends the latter vector back. The result of the inverse F-transform is a function $\hat{f}$ that approximates f.

The original definition of F-transform computes the components using the weighted average which is a simple consequence of minimizing the expression

$$||f(x) - c||_{A_k}$$

over all real constants c, where $|| \cdot ||_{A_k}$ is a norm introduced in a specific space. Therefore, it was not surprising that a generalization of F-transform towards higher degree polynomials (c is a polynomial of zero degree) has appeared relatively early in [14]. Further development including two dimensional case and applications can be found in [15], [16],[17]. In these papers, the authors were interested in the approximation of continuous function using polynomials keeping the same idea as in the original F-transform. Note that all constructions were done in L_2 space and a need to deal with discrete functions was satisfied using not precisely introduced formulas obtained using a discretization of the continues counterparts. The lack of a precise formulation of F-transform of higher degree for discrete functions motivates us to write this contribution that should serve as a starting point for a deeper analysis including statistical one.

For the simplicity, we restrict our development of discrete F^m-transform to so-called generalized uniform fuzzy partitions, nevertheless, one could simply reformulate the definitions and some statements assuming also non-uniform fuzzy partitions.

The structure of this paper is as follows: Section II introduces the concept of generalized uniform fuzzy partition. Section III is devoted to the direct F^m-transform including an analysis of basic properties. The inverse F^m-transform is introduced and some of their properties are derived and discussed in Section IV. Section V concludes the paper.

II. Generalized Uniform Fuzzy Partition

Let $\mathbb{Z}$ and $\mathbb{R}$ denote the set of integers and reals, respectively. It is well-known that a uniform fuzzy partition can be defined using a generating function K which may be modified by a parameter h specifying the required bandwidth. Each basic function of the uniform fuzzy partition is then obtained by a shift of the modified generating function K, where the uniformity for all shifts is supposed. In this paper, the generating function is defined as follows.

Definition 1: A function $K : \mathbb{R} \to [0, 1]$ is said to be a *generating function* if K is an even Lebesgue integrable function (fuzzy set) which is non-increasing in $[0, \infty)$ and

$$K(x) \begin{cases} > 0, & \text{if } x \in (-1, 1); \\ = 0, & \text{otherwise.} \end{cases} \tag{1}$$

A generating function K is said to be *normal* if $K(0) = 1$.

Remark 1: Let us stress that our definition of generating function does not suppose its normality, i.e., $K(0) \in (0, 1]$ in general. Moreover, the continuity of generating function is here naturally replaced by its integrability (cf., [18]).

Example 2 (Triangular generating function): A triangular shaped generating function is a function $K_T : \mathbb{R} \to [0, 1]$ defined by

$$K_T(x) = \max(1 - |x|, 0) \tag{2}$$

for any $x \in \mathbb{R}$.

Example 3 (Raised cosine generating function): A raised cosine generating function is a function $K_C : \mathbb{R} \to [0, 1]$ defined by

$$K_C(x) = \begin{cases} \frac{1}{2}(1 + \cos(\pi x)), & -1 \le x \le 1; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Michal Holčapek is with the University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, Centre of Excellence IT4Innovations, 30. dubna 22, 701 03 Ostrava, Czech Republic (email: Michal.Holcapek@osu.cz)

Tomáš Tichý is with the Technical University of Ostrava, Department of Finance, Centre of Excellence IT4Innovations, Sokolská 33, 701 21 Ostrava, Czech Republic (email: Tomas.Tichy@vsb.cz).

This work has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme "Research and Development for Innovations" funded by Structural Funds of the European Union and state budget of the Czech Republic. The research has been also supported by the Czech Science Foundation through the project of No. 13-13142S.

for any $x \in \mathbb{R}$.

In [19] (see also [20],[21]), a generalization leading to denser (uniform) fuzzy partitions of real intervals with more than two overlapping basic functions has been proposed. In [22], we suggested the following definition of uniform fuzzy partitions of real line (cf., [23]) that generalizes all recent approaches to uniformly defined fuzzy partitions used for fuzzy transform.

Definition 2: Let K be a generating function, h and r be positive real numbers and $x_0 \in \mathbb{R}$. A system of fuzzy sets $\{A_k \mid k \in \mathbb{Z}\}$ defined by

$$A_k(x) = K\left(\frac{x - c_0 - k\,r}{h}\right) \tag{4}$$

for any $k \in \mathbb{Z}$ is called a *generalized uniform fuzzy partition* (GUFP) of the real line determined by the quadruplet (K, h, r, c_0) if the Ruspini condition is satisfied:

$$S(x) = \sum_{k \in \mathbb{Z}} A_k(x) = 1 \tag{5}$$

holds for any $x \in \mathbb{R}$.

In the sequel, the parameters h, r and c_0 are called *bandwidth*, *shift* and *central node*, respectively. The fuzzy sets A_k in (4) that form a uniform fuzzy partition of the real line are called *basic functions*.

A simple consequence of (4) is the formula $A_k(x) = A_0(x - kr)$ that holds for any $x \in \mathbb{R}$ and $k \in \mathbb{Z}$. Putting $c_k = c_0 + kr$ one can simply check that $A_k(c_k)$ is the maximal function value and A_k is centered around the node c_k.

The following theorem provides an equivalent condition to Ruspini's one that can help to verify that a quadruplet (K, h, r, c_0) determines a GUFP. Put $K_h(x) = K(\frac{x}{h})$.

Theorem 1: A quadruplet (K, h, r, c_0) determines a generalized uniform fuzzy partition iff

$$\sum_{i=1}^{\infty} \int_{ir-y}^{y+(i-1)r} K_h(x)dx = y - \frac{r}{2} \tag{6}$$

holds for any $y \in [\frac{r}{2}, r]$.

Proof: It can be found in [22]. ■

Let K be a normal generating function, i.e., $K(0) = 1$. If $\alpha \in (0, 1]$, then we can define the product of the scalar α and the function K by

$$(\alpha K)(x) = \alpha \cdot K(x).$$

Note that in this way, we can determine a family of similarly shaped generating functions and the respective GUFPs, for example, the triangular (raised cosine) shaped generating functions and the triangular (raised cosine) generalized uniform fuzzy partitions.

As a consequence of the preceding theorem we obtain a necessary condition on the GUFP saying that a quadruplet (K, h, r, c_0) to determine a GUFPs has to satisfy the following integral equality:

$$\int_{-1}^{1} K(x)dx = \frac{r}{h}.$$

Hence, considering GUFPs determined by similarly shaped generating functions, we can restrict ourselves to quadruplets $(\frac{r}{h}K, h, r, c_0)$, where K is a normal generating function.

Using the previous theorem, it is not so hard to prove the following useful theorem showing that the necessary and sufficient condition for triangular and raised cosine generalized uniform fuzzy partitions is as simple as possible.

Theorem 2: Let K be the triangular or raised cosine normal generating function. Then, $(\frac{r}{h}K_T, h, r, x_0)$ determines a GUFP iff $\frac{h}{r} \in \mathbb{N}$.

Proof: It can be found in [22]. ■

Note that it is not clear if the previous theorem remains true for an arbitrary normal generating function (e.g., based on B-splines or Shepard kernel [24]). In Figure 1, the triangular and raised cosine generalized uniform fuzzy partitions determined by $(K_T, 2, 0.5, 1)$ and $(0.5 \odot K_C, 2, 1, 1)$ are depicted, respectively.

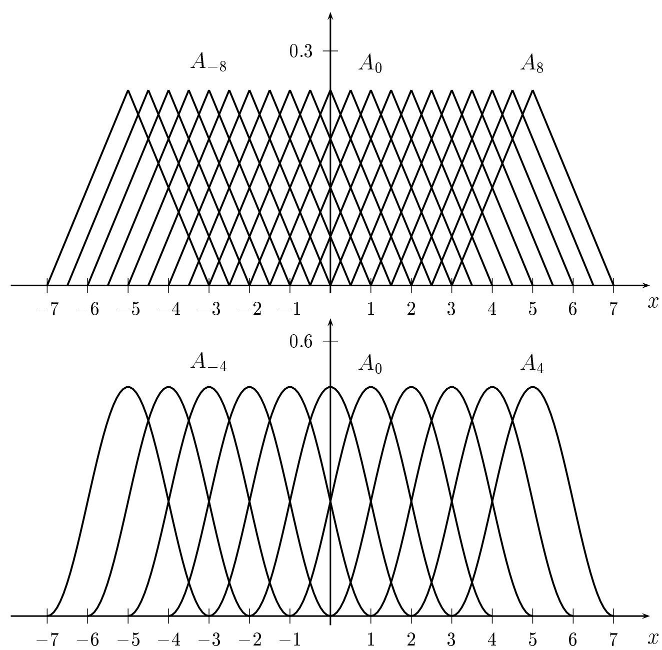

Fig. 1. A part of the triangular and raised cosine GUFP of the real line

III. Direct F^m-transform

A. Preliminaries

Before we provide the definition of direct discrete F^m-transform, let us introduce several preliminary assumptions and notations.

In the sequel, we assume that a discrete function f is given at points $x_1, \ldots, x_n$ (consider $x_i < x_{i+1}$, $i = 1, \ldots, n-1$). Let $\mathbb{A} = \{A_k \mid k \in \mathbb{Z}\}$ be a generalized uniform fuzzy partition, $c_k = c_0 + kr$ denote the k-th node and m refers to the degree of polynomials. Denote by $\mathbf{X} = (x_1, \dots, x_n)^T$ the vector of all values over which the function f is defined, $\mathbf{Y} = (f(x_1), \dots, f(x_n))^T$ the vector of function values of f,

$$\mathbf{X}_k^m = \begin{pmatrix} 1 & x_1 - c_k & \cdots & (x_1 - c_k)^m \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n - c_k & \cdots & (x_n - c_k)^m \end{pmatrix}$$

is the $n \times (m+1)$ Vandermonde matrix and

$$\mathbf{A}_k = \operatorname{diag}\{A_k(x_1), \cdots, A_k(x_n)\}$$

the diagonal matrix of weights computed as the values of basic function A_k at points $x_1, \dots, x_n$. For the sake of simplicity, we omit m in $\mathbf{X}_k^m$ and we write only $\mathbf{X}_k$ if no confusion can appear.

Further, let us use $D(f)$ to denote the domain of f. Since f is discrete, it is easy to see that not all matrices ($k \in \mathbb{Z}$)

$$\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k \tag{7}$$

need to be invertible. Therefore, we restrict ourselves to a subfamily $\mathbb{B} \subset \mathbb{A}$ such that each matrix in the form (7) constructed from a basic function of $\mathbb{B}$ is invertible. This motivates us to introduce the following concept of *sufficiently dense* domain of a discrete function f in a generalized uniform fuzzy partition $\mathbb{A}$ with respect to m.

Definition 3: Let f be a discrete function given at points $x_1, \dots, x_n$, $\mathbb{A}$ be a generalized uniform fuzzy partition and m be a natural number (including 0). The domain of f (or f, for short) is said to be *sufficiently dense* in $\mathbb{A}$ with respect to m if there exists a subfamily of $\mathbb{A}$ denoted by $\mathbb{A}_f$ satisfying the following conditions:

(i) for each $A_k \in \mathbb{A}_f$ the matrix $\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k$ is invertible,
(ii) if $A_{k-1}, A_{k+1} \in \mathbb{A}_f$, then $A_k \in \mathbb{A}_f$,
(iii) $\mathbb{A}_f$ is the largest subfamily of $\mathbb{A}$ satisfying the previous two conditions, i.e., if $\mathbb{B} \subset \mathbb{A}$ satisfies (i) and (ii), then $\mathbb{B} \subseteq \mathbb{A}_f$.

Note that a condition ensuring the sufficient density of f in $\mathbb{A}$ is the existence of a sequence of basic functions $A_k, A_{k+1}, \dots, A_{k+p} \in \mathbb{A}$ such that

$$\#\{x \mid x \in D(f) \text{ and } A_j(x) > 0\} \geq m + 1$$

holds for $j = k, k+1, \dots, k+p$, where $\#\{\dots\}$ denotes the number of elements of a set.[1]

Finally, let use denote by $\mathbf{e}_1$ the unit vector of $m+1$ components having 1 in the first component and zero, elsewhere.

B. Definition

Let f be a discrete function given at the points $x_1, \dots, x_n$ which is sufficiently dense in a generalized uniform fuzzy partition $\mathbb{A}$. Without loss of generality, we always suppose that $\mathbb{A}_f = \{A_1, \dots, A_\ell\}$ for a suitable natural number ℓ.

[1]This follows from the known fact that to construct uniquely a polynomial of degree m which fits the data with respect to the weighted least square error, we need to have at least $m+1$ distinct data.

In the next part, we introduce the discrete direct F^m-transform of two types. The first type of F^m-transform of the m-th degree assigns to f a vector of polynomials of m-th degree in the form

$$\beta_0^k + \beta_1^k(x - c_k) + \dots + \beta_m^k(x - c_k)^m,$$

$k = 1, \dots, \ell$, which locally fits in "a best way" the function f. The second type of F^m-transform of the m-th degree assigns to f a vector of real numbers which corresponds to the values of polynomials obtained in the previous type and evaluated at the points $c_1, \dots, c_k$. It means that the second type of the high degree fuzzy transform is a simplification of the first one, where the polynomials are replaced here by the respective function values. It should be noted that the first type of F^m-transform is a discrete version of the higher degree fuzzy transform defined for continuous functions (cf., [14]), on the other hand, the motivation for the second type of F^m-transform comes from the techniques used in the local polynomial regression (see [25], [26]).

Definition 4 (F^m-transform of type I): Let $\mathbb{A}$ be a GUFP and f be a discrete function given at points $x_1, \dots, x_n$ which is sufficiently dense in $\mathbb{A}$ and denote $\mathbb{A}_f = \{A_1, \dots, A_\ell\}$. A vector of polynomials $(F_1^m[f](x), \dots, F_\ell^m[f](x))$ in the form

$$F_k^m[f](x) = \beta_0^k + \beta_1^k(x - c_k) + \dots + \beta_m^k(x - c_k)^m, \tag{8}$$

where $k = 1, \dots, \ell$, is called *F^m-transform of type I of f with respect to* $\mathbb{A}$ if

$$\boldsymbol{\beta}^k = (\beta_0^k, \dots, \beta_m^k)^T = (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{A}_k \mathbf{Y} \tag{9}$$

for any $k = 1, \dots, \ell$. The polynomial $F_k^m[f]$ is called the *k-th component of F^m-transform of type I of f with respect to* $\mathbb{A}$.

In what follows, if $\mathbf{X} = (x_1, \dots, x_n)^T$, we use

$$F_k^m[f](\mathbf{X}) = \mathbf{X}_k \boldsymbol{\beta}^k \tag{10}$$

to denote the evaluation of the polynomial $F_k^m[f](x)$ at all points $x_1, \dots, x_n$. Note that $F_k^m[f](\mathbf{X})$ is the $n \times 1$ vector, where the j-th row contains precisely the value $F_k^m[f](x_j)$.

The second type of F^m-transform replaces the polynomially expressed components by their specific values as is defined below.

Definition 5 (F^m-transform of type II): Let $\mathbb{A}$ be a GUFP and f be a discrete function given at points $x_1, \dots, x_n$ which is sufficiently dense in $\mathbb{A}$ and denote $\mathbb{A}_f = \{A_1, \dots, A_\ell\}$. A vector of real numbers $(F_1[f], \dots, F_\ell[f])$ is called *F^m-transform of type II of f with respect to* $\mathbb{A}$ if

$$F_k[f] = \mathbf{e}_1 \boldsymbol{\beta}^{\boldsymbol{k}} \tag{11}$$

for any $k = 1, \dots, \ell$, where $\boldsymbol{\beta}^k$ is derived by (9). The real number $F_k[f]$ is called the *k-th component of F^m-transform of type II of f with respect to* $\mathbb{A}$.

In what follows, for the sake of simplicity, we usually omit "with respect to $\mathbb{A}$" in "F^m-transform of type I (II) of f with respect to $\mathbb{A}$", if no confusion can appear. Moreover, sometimes we use F_I^m or F_{II}^m to denote the type I or II of F^m-transform, respectively. It is easy to see that if the vector $(F_1^m[f](x),\dots,F_\ell^m[f])(x)$ is an F_I^m-transform of f, then

$$(F_1^m[f](c_1),\dots,F_\ell^m[f](c_\ell))$$

is the F_{II}^m-transform of f. Obviously, both types coincide in the fuzzy transform of zero degree.

C. Basic properties of F^m-transform components

One can be surprised by a specific form of coefficients of polynomials used in the definition of F^m-transform of type I. The explanation gives the following theorem.

Theorem 3: Let $\boldsymbol{\beta}^k$ be the vector defined in (9). Then, $\boldsymbol{\beta}^k$ minimizes the functional

$$\Phi(b_0,\dots,b_m)=\sum_{i=1}^{n}(f(x_i)-p(x_i))^2A_k(x_i),$$

where $p(x)=b_0+b_1(x-c_k)+\cdots b_m(x-c_k)^m$.

Proof: Using the partial derivatives of the functional $\Phi(b_0,\dots,b_m)$ with respect to $b_0,\dots,b_m$, one can simply derive a system of $m+1$ linear equations with $b_0,\dots,b_m$ as variables which solution minimizes the functional Φ. In the matrix form, the system can be written as follows

$$\mathbf{X}_k^T\mathbf{A}_k\mathbf{X}_k\mathbf{b}=\mathbf{X}_k^T\mathbf{A}_k\mathbf{Y},$$

where $\mathbf{b}=(b_0,\dots,b_m)^T$. Since we assume only such functions that are sufficiently dense with respect to $\mathbb{A}$, the matrix $\mathbf{X}_k^T\mathbf{A}_k\mathbf{X}_k$ is invertible and the solution of the system of linear equations is unique and can be found as

$$\mathbf{b}=(\mathbf{X}_k^T\mathbf{A}_k\mathbf{X}_k)^{-1}\mathbf{X}_k^T\mathbf{A}_k\mathbf{Y}.$$

Hence, we obtain $\mathbf{b}=\boldsymbol{\beta}^k$. ∎

The previous theorem shows that the polynomials as the components of F_I^m-transform provide the best fitting of the values of discrete function f with respect to the weighted least square error. One can see that our definition of discrete F_I^m-transform follows the original idea presented in [1] for zero degree and then in [14] for higher degrees where only the integral weighted least square error is now replaced by its discrete version. The closeness of discrete and continuous direction is, moreover, underline in the following theorem (cf. Lemma 1 in [14]).

First, however, let us define two *column spaces* by $\mathcal{R}(\mathbf{X}_k^m)=\{\mathbf{X}_k^m\mathbf{b}\mid\mathbf{b}^T\in\mathbb{R}^{m+1}\}$ and $\mathcal{R}^n=\{\mathbf{Y}\mid\mathbf{Y}^T\in\mathbb{R}^n\}$ and the scalar product on $\mathcal{R}^n$ by

$$\langle\mathbf{X},\mathbf{Y}\rangle_{\mathbf{A}_k}=\mathbf{X}^T\mathbf{A}_k\mathbf{Y} \tag{12}$$

for any $\mathbf{X},\mathbf{Y}\in\mathcal{R}^n$. Obviously, $\mathcal{R}(\mathbf{X}_k^m)$ is a subspace of the space $\mathcal{R}^n$. The norm on $\mathcal{R}^n$ can be defined by

$$||\mathbf{Y}||_{\mathbf{A}_k}=\langle\mathbf{Y},\mathbf{Y}\rangle_{\mathbf{A}_k}^{1/2}=(\mathbf{Y}^T\mathbf{A}_k\mathbf{Y})^{1/2}. \tag{13}$$

Theorem 4: Let $\mathcal{R}^n$ be the column space enriched with the scalar product $\langle\cdot,\cdot\rangle_{\mathbf{A}_k}$ and $\mathbf{Y}\in\mathcal{R}^n$. The minimum of $||\mathbf{Y}-\boldsymbol{\theta}||_{\mathbf{A}_k}$ for $\boldsymbol{\theta}\in\mathcal{R}(\mathbf{X}_k^m)$ is attained at $\hat{\boldsymbol{\theta}}$ such that $(\mathbf{Y}-\hat{\boldsymbol{\theta}})\perp\mathcal{R}(\mathbf{X}_k^m)$, i.e., when $\mathbf{Y}-\hat{\boldsymbol{\theta}}$ is orthogonal to all vectors in $\mathcal{R}(\mathbf{X}_k^m)$. Moreover, we have

$$\hat{\boldsymbol{\theta}}=\mathbf{X}_k\boldsymbol{\beta}^k,$$

where $\boldsymbol{\beta}^k$ is defined by (9).

Proof: Put $\mathbf{X}_k=\mathbf{X}_k^m$. Let $\hat{\boldsymbol{\theta}}\in\mathcal{R}(\mathbf{X}_k)$ such that $(\mathbf{Y}-\hat{\boldsymbol{\theta}})\perp\mathcal{R}(\mathbf{X}_k)$ and $\boldsymbol{\theta}\in\mathcal{R}(\mathbf{X}_k)$ be arbitrary. Then, we have

$$\begin{aligned}||\mathbf{Y}-\boldsymbol{\theta}||_{\mathbf{A}_k}^2&=(\mathbf{Y}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\mathbf{A}_k(\mathbf{Y}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\&=(\mathbf{Y}-\hat{\boldsymbol{\theta}})^T\mathbf{A}_k(\mathbf{Y}-\hat{\boldsymbol{\theta}})+(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\mathbf{A}_k(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\&\geq||\mathbf{Y}-\hat{\boldsymbol{\theta}}||_{\mathbf{A}_k}^2,\end{aligned}$$

since $(\mathbf{Y}-\hat{\boldsymbol{\theta}})^T\mathbf{A}_k(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ vanishes using the orthogonality assumption. Obviously, the minimum is attained for $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$. Put $\hat{\boldsymbol{\theta}}=\mathbf{X}_k\hat{\mathbf{b}}$. Using the orthogonality, we have $\mathbf{X}_k^T\mathbf{A}_k(\mathbf{Y}-\mathbf{X}_k\hat{\mathbf{b}})=(0,\dots,0)^T$ (consider $\mathbf{X}\mathbf{e}_j\in\mathcal{R}(\mathbf{X}_k)$, $j=1,\dots,m+1$, where $\mathbf{e}_j$ is the unit vector having 1 in its j-th components and zero, elsewhere). Hence,

$$\mathbf{X}_k^T\mathbf{A}_k\mathbf{X}_k\hat{\mathbf{b}}=\mathbf{X}_k^T\mathbf{A}_k\mathbf{Y}.$$

Since we assume that $\mathbf{X}^T\mathbf{A}_k\mathbf{X}_k$ is invertible, we obtain

$$\hat{\boldsymbol{\theta}}=\mathbf{X}_k\hat{\mathbf{b}}=\mathbf{X}_k(\mathbf{X}_k^T\mathbf{A}_k\mathbf{X}_k)^{-1}\mathbf{X}_k^T\mathbf{A}_k\mathbf{Y}=\mathbf{X}_k\boldsymbol{\beta}^{\boldsymbol{k}},$$

which concludes the proof. ∎

A simple consequence of the preceding theorem is the following corollary saying that the quality of approximation using the F^m-transform of type I increases with the increasing degree of polynomial (cf., Lemma 2 in [14]).

Corollary 5: Let f be a discrete function given at the points $x_1,\dots,x_n$ sufficiently dense in $\mathbb{A}$ and $F_k^m[f]$ and $F_k^{m+1}[f]$ be the k-th components of F^m-transform of type I. Then, we have

$$||\mathbf{Y}-F_k^m[f](\mathbf{X})||_{\mathbf{A}_k}\geq||\mathbf{Y}-F_k^{m+1}[f](\mathbf{X})||_{\mathbf{A}_k}, \tag{14}$$

where $\mathbf{X}=(x_1,\dots,x_n)^T$.

Proof: Recall that

$$F_k^m[f](\mathbf{X})=\mathbf{X}_k^m\boldsymbol{\beta}^k \text{ and } F_k^{m+1}[f](\mathbf{X})=\mathbf{X}_k^{m+1}\boldsymbol{\beta'}^k,$$

where $\boldsymbol{\beta}^k$ is the $(m+1)\times 1$ vector minimizing $||\mathbf{Y}-\mathbf{X}_k^m\mathbf{b}||_{\mathbf{A}_k}$ over all $\mathbf{b}$ such that $\mathbf{b}^T\in\mathbb{R}^{m+1}$ and similarly $\boldsymbol{\beta'}^k$ is the respective $(m+2)\times 1$ vector. Since $\mathcal{R}(\mathbf{X}_k^m)\subseteq\mathcal{R}(\mathbf{X}_k^{m+1})\subseteq\mathcal{R}^n$, using the previous theorem, we simply obtain

$$\begin{aligned}||\mathbf{Y}-F_k^m[f](\mathbf{X})||_{\mathbf{A}_k}=||\mathbf{Y}-\mathbf{X}_k^m\boldsymbol{\beta}^k||_{\mathbf{A}_k}\geq\\||\mathbf{Y}-\mathbf{X}_k^{m+1}\boldsymbol{\beta'}^k||_{\mathbf{A}_k}=||\mathbf{Y}-F_k^{m+1}[f](\mathbf{X})||_{\mathbf{A}_k}\end{aligned}$$

and (14) is proved. ∎

Remark 4: Note that the same argument cannot be used in the case of F^m-transform of type II, because, for example, it holds $||\mathbf{Y}-F_k^0(c_k)\mathbf{X}_k^0||_{\mathbf{A}_k}\leq||\mathbf{Y}-F_k^m(c_k)\mathbf{X}_k^0||_{\mathbf{A}_k}$ where recall that $\mathbf{X}_k^0$ is the $n\times 1$ vector having all components equal to 1. The inequality immediately follows from the fact that $\mathbf{Y}-F_k^0(c_k)\mathbf{X}_k^0$ is orthogonal to all vectors of $\mathcal{R}(\mathbf{X}_k^0)$, therefore, $||\mathbf{Y} - F_k^0(c_k)\mathbf{X}_k^0||_{\mathbf{A}_k}$ specifies the shortest distance of $\mathbf{Y}$ from the space determined by $\mathbf{X}_k^0$.

The following lemma shows the linearity of F^m-transform components of type I. As a simple consequence, we obtain also the linearity for type II.

Lemma 6: Let f and g be discrete functions given at the points $x_1, \ldots, x_n$ and $c \in \mathbb{R}$. Then, we have

$$\begin{aligned} F_k^m[f+g](x) &= F_k^m[f](x) + F_k^m[g](x) \\ F_k^m[cf](x) &= cF_k^m[f](x). \end{aligned}$$

Proof: Let f, g be discrete functions given at the points $x_1, \ldots, x_n$ sufficiently dense in $\mathbb{A}$ and denote $\mathbf{Y}_f = (f(x_1), \ldots, f(x_n))$, $\mathbf{Y}_g = (g(x_1), \ldots, g(x_n))$ and $\mathbf{Y}_{f+g} = \mathbf{Y}_f + \mathbf{Y}_g$ (the common sum of vectors). By (9), we have

$$\begin{gathered} F_k^m[f+g] = (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k \mathbf{A}_k \mathbf{Y}_{f+g} = \\ (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k \mathbf{A}_k (\mathbf{Y}_f + \mathbf{Y}_g) = \\ (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k \mathbf{A}_k \mathbf{Y}_f + (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k \mathbf{A}_k \mathbf{Y}_g = \\ F_k^m[f] + F_k^m[g], \end{gathered}$$

where we use the distributivity of the matrix operations. Similarly one can prove the statement for the product of the scalar c and the function f. ∎

IV. Inverse F^m-transform

A. Definition

The inverse F^m-transform for both types is defined as the linear combination of components and basic functions to obtain a continuous function. Note that the original definition of inverse discrete F^m-transform considers a discrete function as the result. The following definition is a slight modification of the original definition introduced in [1] (see, also [18]).

Definition 6: Let $\mathbb{A}$ be a GUFP and f be a discrete function given at points $x_1, \ldots, x_n$ which is sufficiently dense in $\mathbb{A}$ and $\mathbb{A}_f = \{A_1, \ldots, A_\ell\}$. The inverse F^m-transform of type I is a continuous function defined for any $x \in [x_1, x_n]$ by

$$F^m[f](x) = \sum_{k=1}^{\ell} F_k^m[f](x) A_k(x), \tag{15}$$

where $(F_1^m[f](x), \ldots, F_\ell^m[f](x))$ is the direct F^m-transform of type I of f with respect to $\mathbb{A}$.

Definition 7: Let $\mathbb{A}$ be a GUFP and f be a discrete function given at points $x_1 < \cdots < x_n$ which is sufficiently dense in $\mathbb{A}$ and $\mathbb{A}_f = \{A_1, \ldots, A_\ell\}$. The inverse F^m-transform of type II is a continuous function defined for any $x \in [x_1, x_n]$ by

$$F^m[f](x) = \sum_{k=1}^{\ell} F_k^m[f] A_k(x), \tag{16}$$

where $(F_1^m[f], \ldots, F_\ell^m[f])$ is the direct F^m-transform of type II of f with respect to $\mathbb{A}$.

B. Illustration of inverse F^m-transform on financial quantities

In this part, we provide several comparisons among types and degrees of inverse F^m-transform. For the illustration, we use a time series $f(t)$ of financial quantities with 500 observations, i.e., $t \in \{1, \ldots, 500\}$.

On Figure 2, one can see a comparison of F^1-transform (red curve) and F^3-transform (green curve) of type I of $f(t)$ with respect to the (generalized) uniform fuzzy partition $\mathbb{A}$ determined by the quadruplet $(K_C, 40, 40, 0)$. Recall that K_C is the raised cosine generating function, the bandwidth $h = 40$, the shift $r = 40$ and the central node $c_0 = 0$. Note that we put "generalized" into the brackets, because only two basic functions are overlapped here. It is easy to see that the results stated in Corollary 5 can be checked here also visually, it means, there is no doubts about a better approximation of the time series $f(t)$ in the case of F_I^3-transform.

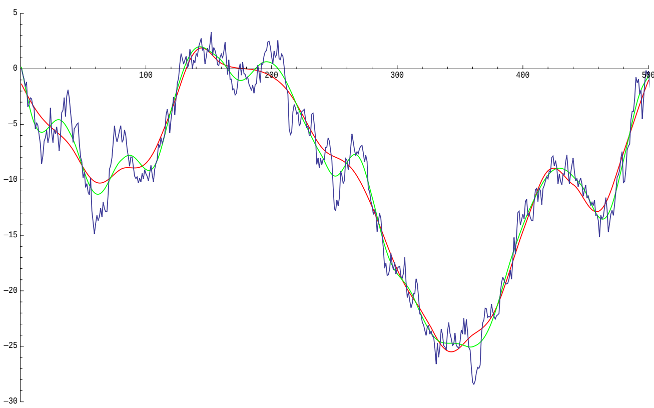

Fig. 2. Comparison of F_I^1-transform (red) and F_I^3-transform (green) of $f(t)$ with respect to the GUFP determined by $(K_C, 40, 40, 0)$

On Figure 3, an alternative comparison of F^1-transform (red curve) and F^3-transform (green curve) of type II of $f(t)$ with respect to $\mathbb{A}$ is depicted. In this case, however, it is not easy to say which resulting function provides a better approximation of the time series $f(t)$.

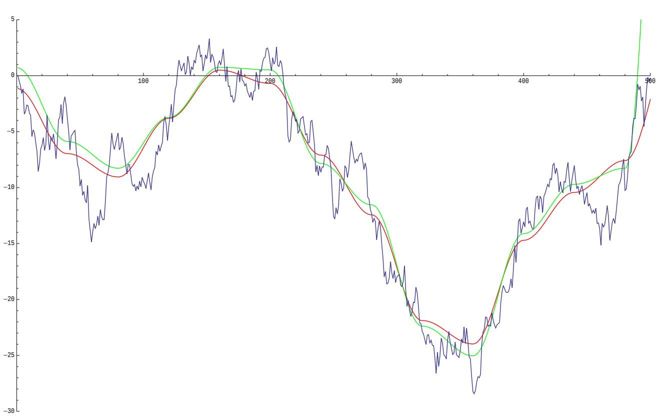

Fig. 3. Comparison of F_{II}^1-transform (red) and F_{II}^3-transform (green) of $f(t)$ with respect to the GUFP determined by $(K_C, 40, 40, 0)$

On Figure 4, one can see that the F^2-transform (green curve) of type II of $f(t)$ provides a smoother resulting function than the F^2-transform (green curve) of type I. For the comparison we used the GUFP determined by $(K_C, 40, 20, 0)$, which is slightly denser than the previous one.

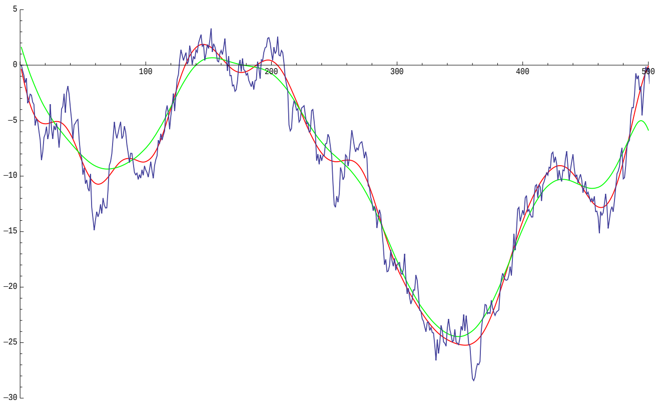

Fig. 4. Comparison of F_I^2-transform (red) and F_{II}^2-transform (green) of $f(t)$ with respect to the GUFP determined by $(K_C, 40, 20, 0)$

On figure 5, one can see a comparison of F^0-transform (red curve) and F^2-transform (green curve) of $f(t)$ of type II. Although, it is not obvious from the figure which resulting function approximates better the time series $f(t)$, according to Remark 4, we know that a better approximation is ensured by the F^0-transform of type II. This fact can be checked

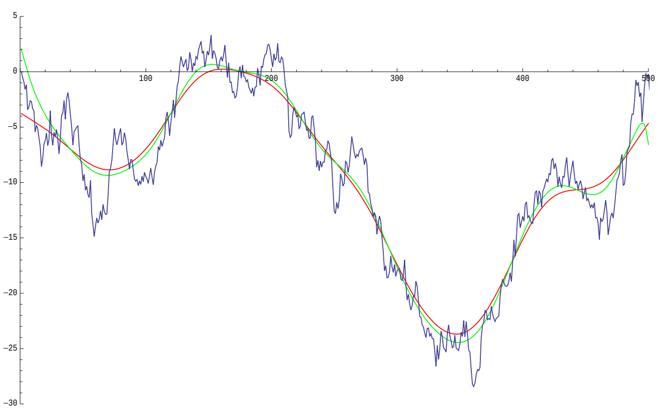

Fig. 5. Comparison of F_{II}^0-transform (red) and F_{II}^2-transform (green) of $f(t)$ with respect to GUFP determined by $(K_C, 40, 4, 0)$

also numerically using the error expressed in the form $||\mathbf{Y} - F_k^m(c_k)\mathbf{X}_k^0||_{\mathbf{A}_k}$ for the k-th component as it is depicted for the first five components in Table I.

C. Basic properties of inverse F^m-transform

In the previous section, we defined the direct F^m-transform of type I and II and showed several basic properties for their components. Now, we are interested in basic properties of resulting functions after the application of inverse F^m-transform. In the sequel, we assume that a GUFP $\mathbb{A}$ is fixed and all discrete functions are sufficiently dense in $\mathbb{A}$.

TABLE I
ERRORS OF APPROXIMATION FOR THE FIRST FIVE COMPONENTS

k	$\lVert\mathbf{Y} - F_k^0(c_k)\mathbf{X}_k^0\rVert_{\mathbf{A}_k}$	$\lVert\mathbf{Y} - F_k^2(c_k)\mathbf{X}_k^0\rVert_{\mathbf{A}_k}$
1	37.495	13 854.4
2	40.791	358.324
3	41.579	2068.37
4	41.044	659.774
5	38.896	614.801

1) Linearity: Recall that it holds $F_k^m[af + bg] = aF_k^m[f] + bF_k^m[g]$ for any F^m-transform component for both types. This kind of linearity can be extended to the resulting functions after the application of inverse of F^m-transform.

Theorem 7: Let f and g be discrete functions given at the points $x_1, \ldots, x_n$ sufficiently dense in $\mathbb{A}$ and $a, b \in \mathbb{R}$. Then, we have

$$F^m[af + bg](x) = aF^m[f](x) + bF^m[g](x),$$

where the F^m-transform is of type I and II.

Proof: We will prove the statement for the F^m-transform of type I, similarly one can prove the statement for the F^m-transform of type II. Let f, g be discrete functions given at the points $x_1, \ldots, x_n$ and assume that $\mathbb{A}_f = \{A_1, \ldots, A_\ell\}$. According to Lemma 6, we have

$$F_k^m[af + bg](x) = aF_k^m[f](x) + bF_k^m[g](x)$$

for any $k = 1, \ldots, \ell$. Hence, we obtain

$$F^m[af + bg](x) = \sum_{k=1}^{\ell} F_k^m[af + bg](x)A_k(x) =$$

$$\sum_{k=1}^{\ell}(aF_k^m[f](x) + bF_k^m[g](x))A_k(x) =$$

$$a\sum_{k=1}^{\ell} F_k^m[f](x)A_k(x) + b\sum_{k=1}^{\ell} F_k^m[g](x))A_k(x) =$$

$$aF^m[f](x) + bF^m[g](x)$$

and the proof is finished. ∎

2) Approximation behaviour: To investigate approximation ability of F^m-transform we will use modulus of continuity. It is easy to see that if $\mathbb{A}$ is a generalized uniform fuzzy partition determined by (K, h, r, c_0) and $x \in \mathbb{R}$, then $A_k(x) = 0$ whenever $c_k \notin [x - h, x + h]$. Therefore, we use the following definition of modulus continuity.

Definition 8: Let f be a discrete function given at points $x_1, \ldots, x_n$ and $h > 0$. A *modulus continuity* of the function f with respect to h is given as a function $\omega_h(f, \cdot) : \mathbb{R} \to \mathbb{R}$ defined by

$$\omega_h(f, x) = \sup_{x_i, x_j \in [x-h, x+h]} |f(x_i) - f(x_j)|. \tag{17}$$

Note that if there is no $x_i \in [x - h, x + h]$ then $\omega_h(f, x) = 0$. It is easy to see that the number $\omega_h(f, x)$ characterizes the smoothness or the continuity of the function f in a specific neighborhood of the point x.

In the next part, we restrict our consideration to type I fuzzy transform. Let us notice that, using Theorem 3, we simply obtain

$$\boldsymbol{\beta}^k = (1, 0, \ldots, 0)^T = (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{A}_k \mathbf{Y}$$

for $\mathbf{Y} = (1, \ldots, 1)^T$ being the $(n \times 1)$ vector. Let us put

$$\mathbf{x}_k^m = (1, x - c_k, \ldots, (x - c_k)^m)$$

for an arbitrary $x \in [x_1, x_n]$. Then, we obtain

$$1 = \mathbf{x}_k^m \boldsymbol{\beta}^k = \mathbf{x}_k^m (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{A}_k \mathbf{Y}.$$

Hence, using the associativity of the matrix product and putting

$$(\alpha_{k1}(x), \ldots, \alpha_{kn}(x)) = \mathbf{x}_k^m (\mathbf{X}_k^T \mathbf{A}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{A}_k,$$

one can derive

$$\alpha_{k1}(x) + \cdots + \alpha_{kn}(x) = 1. \tag{18}$$

Put $\boldsymbol{\alpha}_k(x) = (\alpha_{k1}(x), \ldots, \alpha_{kn}(x))$. Then, we can interpret the value $F_k^m[f](x)$ (the value of polynomial attained at the point x) as a linear combination of function values of f at the points $x_1, \ldots, x_n$ and the coefficients of the vector $\boldsymbol{\alpha}_k(x)$, i.e.,

$$F_k^m[f](x) = \boldsymbol{\alpha}_k(x)(f(x_1), \ldots, f(x_n))^T = \\ \alpha_{1k}(x)f(x_1) + \cdots + \alpha_{nk}(x)f(x_n).$$

Unfortunately, it is not clear to us if $\alpha_{kj}(x) \geq 0$ for any $k = 1, \ldots, \ell$ and $j = 0, \ldots, m$ for any $x \in [x_1, x_n]$ or at least for any $x = x_1, \ldots, x_n$ holds in general and we leave the answer to this question to our future research. Nevertheless, assuming for the moment the non-negativity of all coefficients the $n \times 1$ vector $\boldsymbol{\alpha}_k(x)$ can be interpreted as a vector of weights and each component of F^m-transform evaluated at a point x can be obtained as a weighted average of function values of f.

Theorem 8: Let $\mathbb{A}$ be a GUFP detemined by (K, h, r, c_0), f be a discrete function given at points $x_1, \ldots, x_n$ and put $\mathbb{A}_f = \{A_1, \ldots, A_\ell\}$. Let us assume that

$$\alpha_{kj}(x_i) \geq 0 \tag{19}$$

for any $i = 1, \ldots, n$, $k = 1, \ldots, \ell$ and $j = 1, \ldots, n$. Then,

$$|f(x_i) - F^m[f](x_i)| \leq \omega_h(f, x_i) \tag{20}$$

for any $i = 1, \ldots, \ell$.

Proof: Let $x_i \in D(f)$. Then,

$$|f(x_i) - F^m[f](x_i)| = \\ |\sum_{k=1}^{\ell} f(x_i)A_k(x_i) - \sum_{k=1}^{\ell} F_k^m[f](x_i)A_k(x_i)| \\ \leq \sum_{k=1}^{\ell} |f(x_i) - F_k^m[f](x_i)|A_k(x_i) = \\ \sum_{k=1}^{\ell} |f(x_i) - \sum_{j=1}^{n} \alpha_{kj}(x_i)f(x_j)|A_k(x_i) \leq \\ \sum_{k=1}^{\ell}\sum_{j=1}^{n} \alpha_{kj}(x_i)|f(x_i) - f(x_j)|A_k(x_i) \leq \\ \sum_{k=1}^{\ell}\sum_{j=1}^{n} \alpha_{kj}(x_i)\omega_h(f, x_i)A_k(x_i) = \omega_h(f, x_i),$$

where we used $\alpha_{k1}(x_i) + \cdots + \alpha_{kn}(x_i) = 1$ and the assumption (20) on the non-negativity of values $\alpha_{kj}(x_i)$. ■

As we have mentioned before, there is a question if the assumption (19) is true for all or some degrees of polynomials used in F^m-transform, or even it depends on data. A solution of this open problem is left to our future research.

V. Conclusions

In this paper, we introduced discrete version of direct and inverse fuzzy transform of higher degree of type I and II. We used matrices to define the direct F^m-transform by means of polynomials and showed that our definition of polynomials leads to a minimization of the standard functional with respect to the weighted least square error. Moreover, we proved that polynomials of higher degrees provides a better quality of approximation of the original function. Finally, we introduced the inverse F^m-transform and derived some basic facts. We illustrated the discrete approach to F^m-transform on financial data.

In future, we want to solve the open problem stated in the previous paragraph concerning the satisfaction of the inequality in (19). Further, we want to investigate statistical properties of discrete F^m-transform, since both types can serve as non-parametric estimators. It is known that the F^m-transform of type II is a discrete version of local polynomial regression, therefore, we can adopt many result from the regression theory. A comparison of the standard "continuous"' approaches and new "discrete" ones seems to be very useful also from the practical point of view, e.g., a reduction of complexity for a manipulation with large data. Finally, a generalization of discrete F^m-transform to multivalued functions is also a topic of our future research.

References

[1] 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.

[2] ——, "Fuzzy transforms: Theory and applications," *Fuzzy sets syst.*, vol. 157(8), pp. 993–1023, 2006.
[3] M. Daňková and R. Valášek, "Full fuzzy transform and the problem of image fusion," *J. Electr. Eng.*, vol. 57, no. 7, pp. 82–84, 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*, vol. 48, no. 1, pp. 110–131, 2008.
[5] F. Di Martino, V. Loia, and S. Sessa, "Fuzzy transforms method and attribute dependency in data analysis," *Inf. Sci.*, vol. 180, no. 4, pp. 493–505, 2010.
[6] F. Di Martino, V. Loia, and S. Sessa, "A segmentation method for images compressed by fuzzy transforms," *Fuzzy Sets Syst.*, vol. 161, no. 1, pp. 56–74, 2010.
[7] F. Di Martino, V. Loia, and S. Sessa, "Fuzzy transforms method in prediction data analysis," *Fuzzy Sets Syst.*, vol. 180, no. 1, pp. 146–163, 2011.
[8] V. Novák, I. Perfilieva, M. Holčapek, and V. Kreinovich, "Filtering out high frequencies using f-transform," *Submitted to Information Sciences*.
[9] 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.
[10] I. Perfilieva, V. Novák, and A. Dvořák, "Fuzzy transform in the analysis of data," *Int. J. Approx. Reasoning*, vol. 48, no. 1, pp. 36–46, 2008.
[11] M. Štěpnička, "Fuzzy transformation and its applications in a/d converter," *J. Electr. Eng.*, vol. 54, no. 12, pp. 72–75, 2003.
[12] M. Štěpnička and O. Polakovič, "A neural network approach to the fuzzy transform," *Fuzzy Sets Syst.*, vol. 160, no. 8, pp. 1037–1047, 2009.
[13] M. Štěpnička and R. Valášek, "Fuzzy transforms and their application to wave equation," *J. Electr. Eng.*, vol. 55, no. 12, pp. 7–10, 2004.
[14] I. Perfilieva and M. Daňková, "Towards F-transform of a higher degree," in *Proceedings of IFSA/EUSFLAT 2009*, Lisbon, Portugal, 2009, pp. 585–588.
[15] I. Perfilieva and V. Kreinovich, "Fuzzy transform of higher order approximate derivatives: A theorem," *Fuzzy sets syst.*, vol. 180, pp. 55–68, 2011.
[16] I. Perfilieva, P. Hodáková, and P. Hurtík, "F^1-transform edge detector inspired by canny's algorithm," in *Communications in Computer and Information Science. Advances on Computational Intelligence*, 2012, pp. 230–239.
[17] I. Perfilieva and P. Hodáková, "F^1-transform of functions of two variables," in *Proc. EUSFLAT 2013, Advances in Intelligent Systems Research*, Milano, Italy, 2013, pp. 547–553.
[18] I. Perfilieva, M. Daňková, and B. Bede, "Towards a higher degree F-transform." *Fuzzy Sets Syst.*, vol. 180, no. 1, pp. 3–19, 2011.
[19] L. Stefanini, "Fuzzy transform and smooth function," in *Proceedings of IFSA/EUSFLAT 2009*, Lisbon, Portugal, 2009, pp. 579–584.
[20] ——, "F-transform with parametric generalized fuzzy partitions," *Fuzzy Sets Syst.*, vol. 180, no. 1, pp. 98–120, 2011.
[21] M. Holčapek and T. Tichý, "A smoothing filter based on fuzzy transform." *Fuzzy Sets Syst.*, vol. 180, no. 1, pp. 69–97, 2011.
[22] M. Holčapek, V. Novák, I. Perfilieva, and V. Kreinovich, "Necessary and sufficient conditions for generalized uniform fuzzy partitions," *Submitted toFuzzy Sets Syst.*, 2014.
[23] K. Loquin and O. Strauss, "Histogram density estimators based upon a fuzzy partition," *Stat. Probab. Lett.*, vol. 78, no. 13, pp. 1863–1868, 2008.
[24] B. Bede and I. J. Rudas, "Approximation properties of fuzzy transforms," *Fuzzy sets syst.*, vol. 180, no. 1, pp. 20–40, 2011.
[25] A. Pagan and A. Ullah, *Nonparametric Econometrics*. New York: Cambridge University Press, 1999.
[26] M. P. Wand and M. Jones, *Kernel Smoothing*. London: Chapman&Hall/CRC Monographs on Statistics & Applied Probability, 1995.