Fuzzy Transform and least-squares approximation: analogies, differences, and generalizations

Giuseppe Patané
CNR-IMATI[1]
patane@ge.imati.cnr.it

Abstract

This paper investigates the relations between the least-squares approximation techniques and the Fuzzy Transform. Assuming that the function $\overline{f}:\mathbb{R}^d \rightarrow \mathbb{R}$ *underlying a discrete data set* $\mathcal{D} := \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^{s}$ *has been computed with interpolating or least-squares constraints, we prove that the Discrete Fuzzy Transform of the sets* $\{f(\mathbf{x}_i)\}_{i=1}^{s}$ *and* $\{\overline{f}(\mathbf{x}_i)\}_{i=1}^{s}$ *is the same. This result shows that the Discrete Fuzzy Transform is invariant with respect to the interpolating and least-squares approximation of* $\mathcal{D}$*. Additionally, the Fuzzy Transform of* f *outside* $\mathcal{P}$ *is approximated by simply resampling the continuous map* $\overline{f}$ *at a set of points of* $\mathbb{R}^d\backslash\mathcal{P}$*. Using numerical linear algebra, we also derive new properties (e.g., stability to noise, additivity with respect to* $\mathcal{P}$*) and characterizations (e.g., radial and dual membership maps) of the Discrete Fuzzy Transform. Finally, we define the geometry- and confidence-driven Discrete Fuzzy Transform, which take into account the intrinsic geometry of the input data and the confidence weights associated to the* f*-values or the points of* $\mathcal{P}$*.*

Keywords: Fuzzy Transform, Discrete Fuzzy Transform, least-squares approximations, radial basis functions, dual basis, Laplacian matrix, intrinsic geometry.

1. Introduction

During the last decades, several transformations have been proposed to solve problems which spread from signal analysis to the solution of differential equations and the approximation of scalar functions. Among them, we mention the Fourier and wavelet transform in signal and image processing, linear transformations associated to the Laplace and the heat diffusion operator in differential analysis. In all the aforementioned cases, the underlying idea is to compute an approximation of the solution as member of a linear space, which is generated by a set of basis functions and whose properties (e.g., continuity, reproducing

[1]Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Genova, Italy

property in Hilbert spaces) are strictly related to the input data and problem. For instance, the basis functions of the Fourier and wavelet transform are sinusoidal, the function space associated to the Laplace and heat diffusion operator is generated by the eigenfunctions of the Laplace-Beltrami operator. We can also consider the wavelet, time-scale and time-frequency basis induced by the Haar and Gabor atoms (Daubechies, 1992).

In fuzzy modeling, the *Fuzzy Transform* $\mathcal{F} : \mathcal{C}^0 \rightarrow \mathbb{R}^n$ provides a relation between the space $\mathcal{C}^0$ of continuous functions defined on the real line and $\mathbb{R}^n$. In a similar way, the *Inverse Fuzzy Transform* $\mathcal{F}_{-1} : \mathbb{R}^n \rightarrow \mathcal{C}^0$ identifies any vector of $\mathbb{R}^n$ with a continuous map. Even though $\mathcal{F}_{-1}$ is not the inverse of $\mathcal{F}$ (i.e., $\mathcal{F} \circ \mathcal{F}_{-1}$ is not the identity functional), under mild conditions on the input f-values the Inverse Fuzzy Transform $f_{F,n}$ approximates f up to an arbitrary precision (Perfilieva, 2006). In particular, discrete input data can be converted to a continuous approximation through the Inverse Fuzzy Transform and computations on $\mathcal{C}^0$ are changed into discrete operations on $\mathbb{R}^d$ through the Fuzzy Transform.

The ubiquity of the Fuzzy Transform is provided by its different constructions, which exploit the linear algebra in vector spaces of finite dimension and the residuated lattice. Among the main applications, we mention data mining (Agrawal et al., 1993; Hong et al., 2003; Mitra et al., 2002; Zhang et al., 2006); knowledge discovery (Fayyad et al., 1996; Piatetsky-Shapiro, 2000); the analysis of linguistic expressions (Novák et al., 2008); the approximation of discrete data (Perfilieva, 2005, 2006; Perfilieva et al., 2008; Stepnicka and Polakovic, 2009); image analysis and compression (Di Martino et al., 2008).

In this context, we investigate the relations between the least-squares approximation of a set of scattered data $\mathcal{D} := \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^{s}$ and the Fuzzy Transform. Here, $\mathcal{P} := \{\mathbf{x}_i\}_{i=1}^{s}$ is a set of points in $\mathbb{R}^d$ and the corresponding f-values are associated to an unknown function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Using least-squares techniques, we compute the function $\overline{f} : \mathbb{R}^d \rightarrow \mathbb{R}$ *underlying* f, which is defined as a smooth map that locally approximates $\mathcal{D}$. Since exactly reproducing all the f-values generally leads to unstable approximations with low generalization properties and to recover the noise component, we focus our investigation on the least-squares framework (Golub and VanLoan, 1989). In order to provide a good approximation accuracy and robustness to noise, the approximation $\overline{f}(\mathbf{x}) := \sum_{i=1}^{n} \alpha_i A_i(\mathbf{x})$ is searched in the linear space generated by the set $\mathcal{A} := \{A_i(\mathbf{x})\}_{i=1}^{n}$, where each map A_i is (i) a membership function of a fuzzy partition of the input data set or (ii) a radial basis function $A_i(\mathbf{x}) := \varphi(\|\mathbf{x} - \mathbf{x}_i\|_2)$ generated by a kernel map $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}$ (Aronszajn, 1950; Dyn et al., 1986; Farwig, 1986; Micchelli, 1986; Schoelkopf and Smola, 2002; Wendland, 1995). The main reason we use the radial membership functions is that they provide a representation of the approximating function which is efficiently computed through the solution of a sparse linear system and is oblivious of the organization, sampling, and connectivity (e.g., real line, regular grid) of the input data.

In the second part of the paper, we study the relations between the approximation $\overline{f} : \mathbb{R}^d \rightarrow \mathbb{R}$, the direct, and the Inverse Fuzzy Transform of $\mathcal{D}$ and $\overline{f}$. Assuming that $\overline{f}$ has been computed with interpolating or least-squares con-

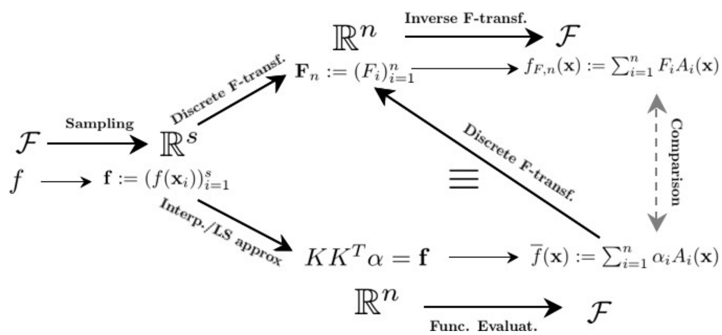

Figure 1: Scheme of the relations between the least-squares and the Fuzzy Transform that will be analyzed throughout the paper.

straints on $\mathcal{D}$, we prove that the Discrete Fuzzy Transform of the two sets $\{f(\mathbf{x}_i)\}_{i=1}^s$ and $\{\overline{f}(\mathbf{x}_i)\}_{i=1}^s$ is the same. This result shows that the Discrete Fuzzy Transform is invariant with respect to the interpolating and least-squares approximation of $\mathcal{D}$. Furthermore, resampling the continuous map $\overline{f}$ at a set of points of $\mathbb{R}^d \backslash \mathcal{P}$ allows us to approximate the Discrete Fuzzy Transform of f outside $\mathcal{P}$. Numerical properties (e.g., stability to noise, additivity with respect to $\mathcal{P}$) and characterizations (e.g., radial and dual membership maps) of the proposed framework will be derived by introducing a matrix formulation of the Discrete Fuzzy Transform. Finally, to better analyze the relation between the Inverse Fuzzy Transform $f_{F,n}$ and the approximation $\overline{f}$, the membership functions are converted in a *dual form* by making the f-values explicit in the expression of $f_{F,n}$ and $\overline{f}$. More precisely, $f_{F,n}$ (or $\overline{f}$) is rewritten in terms of the dual basis $\mathcal{B} := \{B_i(\mathbf{x})\}_{i=1}^s$ as $f_{F,n}(\mathbf{x}) := \sum_{i=1}^s f(\mathbf{x}_i)B_i(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$.

Observing that the Discrete Fuzzy Transform involves the f-values and the membership functions without any information on the distribution or confidence of the input data $\mathcal{D}$, we define the *geometry-* and *confidence-driven Discrete Fuzzy Transform*. These two formulations, which take into account the intrinsic geometry of $\mathcal{P}$ and the confidence weights associated to the f-values, generalize the Discrete Fuzzy Transform while preserving its main properties. We will show that enriching $\mathcal{D}$ with geometric and confidence information, which is extracted with statistical and graph-based techniques, is crucial to improve the approximation accuracy of $\overline{f}$ and the Inverse Fuzzy Transform for structured data (e.g., functions, images, manifold surfaces). For instance, this is the case of the values of a function sampled on a manifold surface, whose dimension is lower than that of the embedding space; the pixels of an image; the points generated by physical processes; or the maps that are solutions of differential equations. Figure 1 summarizes the main steps of the proposed framework.

Paper organization. We briefly introduce previous work on the Fuzzy Transform and its specialization to radial membership functions (Section 2). Then, we study the relation between the Fuzzy Transform and function approximation techniques (Section 3). Using these results, the approximation and Fuzzy

Transform are enriched with geometric and confidence information (Section 4). Finally, limits and possible extensions of the proposed approach are presented (Section 5).

2. Fuzzy Transform: definitions and specialization to radial membership functions

Firstly, we introduce the theoretical background on the Fuzzy Transform and function approximation (Section 2.1); then, we generalize the Fuzzy Transform to the case of radial membership functions (Section 2.2).

2.1. Theoretical background and previous work

We briefly review previous work on the Fuzzy Transform; for more details, we refer the ready to (Perfilieva, 2006; Perfilieva et al., 2008). Let $\mathcal{I} := \{\Omega_i\}_{i=1}^n$ be a partition of a compact set Ω of $\mathbb{R}^d$ and $\mathcal{C} := \{\mathbf{p}_i\}_{i=1}^n$ a set of points such that $\mathbf{p}_i \in \Omega$, $i = 1, \ldots, n$. A family of functions $\mathcal{A} := \{A_i : \Omega \to [0, 1]\}_{i=1}^n$ is a *fuzzy partition* of Ω if the following properties hold for each i

- $A_i(\mathbf{x}) \neq 0$, $\mathbf{x} \in \Omega_i$ and $A_i(\mathbf{p}_i) = 1$;
- A_i is continuous and has its unique maximum at $\mathbf{p}_i$;
- for all $\mathbf{x} \in \Omega$, $\sum_{i=1}^n A_i(\mathbf{x}) = 1$.

Under these assumptions, the *Fuzzy Transform* (Perfilieva, 2006) of a function $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ is defined as the array $\mathbf{F}_n := (F_k)_{k=1}^n \in \mathbb{R}^n$ whose components are

$$F_k := \frac{\int_{\mathbb{R}^d} f(\mathbf{x})A_k(\mathbf{x})d\mathbf{x}}{\int_{\mathbb{R}^d} A_k(\mathbf{x})d\mathbf{x}}, \qquad k = 1, \ldots, n. \tag{1}$$

From now on, we omit the integration domain. Among the properties of the Fuzzy Transform, we mention the linearity with respect to the input function and the least-squares property, which guarantees that the k^{th} component of $\mathbf{F}_n$ minimizes the quadratic least-squares error $\Phi_k(t) := \int |f(\mathbf{x}) - t|^2 A_k(\mathbf{x})d\mathbf{x}$, $t \in \mathbb{R}$, associated to A_k.

In real situations, where the function f is known only at a given set of points $\mathcal{P} := \{\mathbf{x}_i\}_{i=1}^s$, the definition (1) is replaced by the *Discrete Fuzzy Transform* $\mathbf{F}_n := (F_k)_{k=1}^n \in \mathbb{R}^n$, whose components are

$$F_k := \frac{\sum_{j=1}^s f(\mathbf{x}_j)A_k(\mathbf{x}_j)}{\sum_{j=1}^s A_k(\mathbf{x}_j)}, \qquad k = 1, \ldots, n.$$

In this case, the number s of samples is smaller than or equal to the number n of membership functions. The Discrete Fuzzy Transform can be used to recover an approximation $f_{F,n}$ of the function f underlying the set of values $\{f(\mathbf{x}_i)\}_{i=1}^s$ through the *Inverse Fuzzy Transform* (Perfilieva, 2006), which is defined as

$$f_{F,n}(\mathbf{x}) := \sum_{k=1}^n F_k A_k(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^d.$$

Finally, for any given approximation accuracy ϵ there exists a number n_ϵ of nodes and a set $\{A_k(\mathbf{x})\}_{k=1}^{n_\epsilon}$ of n_ϵ membership functions such that the discrepancy $\|f - f_{F,n_\epsilon}\|_\infty$ between f and its Inverse Fuzzy Transform f_{F,n_ϵ} is lower than ϵ.

2.2. Fuzzy Transform using radial membership functions

In the following, we specialize the Fuzzy Transform to the case of radial membership functions and show the relations between this specialization and previous work. In this way, we address the computation of the Fuzzy Transform of functions defined on subsets of $\mathbb{R}^d$, which are not necessarily organized on regular (e.g., rectangular, voxel, tetrahedral) grids. To this end, let us assume that the membership functions $\{A_i(\mathbf{x})\}_{i=1}^n$ of the fuzzy partition are generated by a *kernel* $\varphi : \mathbb{R}^+ \to \mathbb{R}$ as

$$A_i(\mathbf{x}) := \varphi(\|\mathbf{x} - \mathbf{p}_i\|_2), \qquad \mathbf{x} \in \mathbb{R}^d, \qquad i = 1, \dots, n. \tag{2}$$

Indeed, each membership map A_i is centered at $\mathbf{p}_i$, radially symmetric, and generated by φ. Note that several membership functions can be written as Equation (2); for instance, for the Gaussian case the kernel function is $\varphi(t) := \exp(-t/h)$, where h is the kernel support. Other examples include the triangular and sinusoidal shaped basis functions, where φ is given by $\varphi_1(t) := \frac{1-t}{h}$, $\varphi_2(t) := \frac{t}{h}$ and $\varphi(t) := \frac{\cos t}{h}$, $t \in \mathbb{R}$, respectively. We now show that a set of simple constraints on φ guarantees that the corresponding radial basis functions satisfy the properties of the membership maps listed in Section 2.1. More precisely, we have that for any $i = 1, \dots, n$,

- $\varphi(0) = 1$ implies $A_i(\mathbf{p}_i) = 1$;
- if φ is positive/increasing/decreasing, then A_i has the same behaviour;
- if φ has compact support $\text{supp}(\varphi) := \overline{\{t \in \mathbb{R} : \varphi(t) \neq 0\}}$, then A_i has compact support. In fact, if $\text{supp}(\varphi) = [a, b]$ we have that the support of A_i is a closed subset of the compact set $\{\mathbf{x} \in \mathbb{R}^d : a \leq \|\mathbf{x} - \mathbf{p}_i\|_2 \leq b\}$; indeed, the support of A_i is compact in $\mathbb{R}^d$;
- if φ is continuous, then A_i is (at least) continuous;
- the partition of the unity property is achieved by normalizing the membership function $A_i(\mathbf{x})$ as $\tilde{A}_i(\mathbf{x}) := \frac{A_i(\mathbf{x})}{\sum_{j=1}^n A_j(\mathbf{x})}$. In particular, we have that $|\tilde{A}_i(\mathbf{x})| \leq 1$, $\mathbf{x} \in \mathbb{R}^d$, $i = 1, \dots, n$.

Assuming that the kernel φ satisfies the previous properties and that f is continuous, we define the *continuous Fuzzy Transform* of the map $f : \mathbb{R}^d \to \mathbb{R}$ as the function $F : \mathbb{R}^d \to \mathbb{R}$ given by

$$F(\mathbf{x}) := \frac{\int f(\mathbf{y})\varphi(\|\mathbf{x} - \mathbf{y}\|_2)d\mathbf{y}}{\int \varphi(\|\mathbf{x} - \mathbf{y}\|_2)d\mathbf{y}}, \qquad \mathbf{x} \in \mathbb{R}^d. \tag{3}$$

First of all, we verify that the value $F(\mathbf{x})$ in (3) is well-defined; in fact, let us indicate with $\varphi_{\mathbf{x}}(\mathbf{y}) := \varphi(\|\mathbf{x} - \mathbf{y}\|_2)$ the membership function centered at $\mathbf{x}$.

Then, the support $\Sigma_{\mathbf{x}} := \overline{\{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{x}-\mathbf{y}\|_2 \in \mathrm{supp}(\varphi)\}}$ of $\varphi_{\mathbf{x}}$ is compact; from the continuity of f, it follows that for any $\mathbf{x} \in \mathbb{R}^d$ the value $\max_{\mathbf{y}\in\Sigma_{\mathbf{x}}}\{|f(\mathbf{y})|\}$ is finite and $|F(\mathbf{x})| \leq \max_{\mathbf{y}\in\Sigma_{\mathbf{x}}} |f(\mathbf{y})|$ is well-defined.

Since $F(\mathbf{x}_k)$ is the k^{th} component of the Fuzzy Transform of f, the function F interpolates the values $\mathbf{F}_n := (F_k)_{k=1}^n$ of the Discrete Fuzzy Transform associated to the set $\{f(\mathbf{x}_i)\}_{i=1}^s$ of f-values. Finally, F still verifies the least-squares property by minimizing the *energy function* defined as

$$E(\mathbf{y}, t) := \int |f(\mathbf{x}) - t|^2 \varphi(\|\mathbf{x}-\mathbf{y}\|_2) d\mathbf{x}, \qquad \mathbf{y} \in \mathbb{R}^d, \qquad t \in \mathbb{R}.$$

In fact, deriving the function E with respect to t, we have that $\partial_t E(\mathbf{y}, t) = 0$ if and only if $t = F(\mathbf{x})$. The interpolation of the $\mathbf{F}_n$-values and the least-squares property allow us to consider F as a *generalization of the Fuzzy Transform* for radial membership functions.

Let us consider the *Fuzzy operator* defined as

$$\mathcal{F} : \mathcal{C}^0 \to \mathcal{C}^1, \qquad f \mapsto \mathcal{F}(f) := F,$$

where we consider the L_∞-norm on both spaces $\mathcal{C}^0$ and $\mathcal{C}^1$. From the previous definition and the assumption that φ has a compact support, we have that the F-values are bounded by the L_∞-norm of f; in fact,

$$|F(\mathbf{x})| = \left| \frac{\int f(\mathbf{y})\varphi(\|\mathbf{x}-\mathbf{y}\|_2)d\mathbf{y}}{\int \varphi(\|\mathbf{x}-\mathbf{y}\|_2)d\mathbf{y}} \right| \leq \|f\|_\infty, \qquad \mathbf{x} \in \mathbb{R}^d. \tag{4}$$

In particular, F is well-defined at each point $\mathbf{x}$ of $\mathbb{R}^d$. Since $\mathcal{F}$ is the composition of linear operators, it is linear; i.e., $\forall f, g \in \mathcal{C}^0$, $\forall \alpha, \beta \in \mathbb{R}$, the identity $\mathcal{F}(\alpha f + \beta g) = \alpha\mathcal{F}(f) + \beta\mathcal{F}(g)$ holds. The continuity of $\mathcal{F}$ follows from the relation (4) and $\|\mathcal{F}(f)\|_\infty \leq \|f\|_\infty$, $\forall f \in \mathcal{C}^0$.

3. Fuzzy Transform and function approximation

In the previous section, we have provided a continuous function F that interpolates the values of $\mathbf{F}_n$. In the following, we investigate the properties of both the Fuzzy Transform and its discrete version in the context of the approximation theory. To this end, we approximate the input data set $\mathcal{D} := \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^s$ with a continuous representation $\overline{f} : \mathbb{R}^d \to \mathbb{R}$ and prove the invariance of the Discrete Fuzzy Transform of $\overline{f}$ with respect to the interpolating/least-squares approximation of $\mathcal{D}$. As detailed in Section 4, this result also motivates the use of the least-squares approximation as basis for the insertion of additional constraints in the definition of the Fuzzy Transform. Among these constraints, we mention the geometric distribution of the input data or confidence weights associated to the f-values.

To address the aforementioned aims, we introduce a matrix formulation of the Discrte Fuzzy Transform, estimate its energy, and discuss its stability to noise (Section 3.1). Then, we investigate the relations between the Fuzzy Transform and the least-squares approximation (Section 3.2).

3.1. Matrix formulation of the Discrete Fuzzy Transform: energy estimation and stability to noise

To define the matrix formulation of the Discrete Fuzzy Transform, let us introduce the coefficient matrix $K := (A_k(\mathbf{x}_i))_{k=1,\dots,n}^{i=1,\dots,s} \in \mathbb{R}^{n\times s}$, $n \leq s$, and the $n \times n$ diagonal matrix $D := \mathrm{diag}(d_1, \dots, d_n)$ with entries $d_j := \sum_{i=1}^{s} A_j(\mathbf{x}_i)$, $j = 1, \dots, n$. Then, we have that the Discrete Fuzzy Transform $\mathbf{F}_n$ of f, with respect to the set of points $\mathcal{P} := \{\mathbf{x}_i\}_{i=1}^{s}$, is represented in matrix form as

$$\mathbf{F}_n := (F_k)_{k=1}^{n} = D^{-1}K\mathbf{f}, \qquad \mathbf{f} := (f(\mathbf{x}_i))_{i=1}^{s} \in \mathbb{R}^s. \tag{5}$$

From the previous relations, we have that

$$\begin{aligned}
|F_k| &= \frac{|\sum_{i=1}^{s} f(\mathbf{x}_i)A_k(\mathbf{x}_i)|}{\sum_{i=1}^{s} A_k(\mathbf{x}_i)}\\
&\leq \frac{\sum_{i=1}^{s} |f(\mathbf{x}_i)|A_k(\mathbf{x}_i)}{\sum_{i=1}^{s} A_k(\mathbf{x}_i)}\\
&\leq \|\mathbf{f}\|_\infty, \qquad k = 1, \dots, n;
\end{aligned}$$

indeed, $\|\mathbf{F}_n\|_\infty \leq \|\mathbf{f}\|_\infty$ and $\|\mathbf{F}_n\|_2 \leq \sqrt{n}\|\mathbf{f}\|_\infty$. It follows that each value F_k is bounded by the maximum absolute value of f. We now derive an estimation of the *energy* $\|\mathbf{F}_n\|_2$ of the Discrete Fuzzy Transform using its matrix formulation (5); in fact,

$$\begin{aligned}
\|\mathbf{F}_n\|_2 &= \|D^{-1}K\mathbf{f}\|_2\\
&\leq \|D^{-1}\|_2\|K\|_2\|\mathbf{f}\|_2\\
&= \left[\min_{i=1,\dots,n}\left\{\sum_{j=1}^{s} A_i(\mathbf{x}_j)\right\}\right]^{-1} \sigma_{\max}(K)\|\mathbf{f}\|_2,
\end{aligned}$$

where $\sigma_{\max}(K)$ is the maximum singular value of the matrix K. We conclude that the l_2 energy of $\mathbf{F}_n$ is proportional to the energy of the input signal and $\sigma_{\max}(K)$. Finally, we notice that the values of the Inverse Fuzzy transform $f_{F,n}$ at the points of $\mathcal{P}$ can be written as $(f_{F,n}(\mathbf{x}_i))_{i=1}^{s} = K^T D^{-1}K\mathbf{f}$.

Stability of the Discrete Fuzzy Transform to noise. To study the stability of the Discrete Fuzzy Transform with respect to the input data, let us consider the f-values $\mathbf{f} := (f(\mathbf{x}_i))_{i=1}^{s}$ and its perturbation $\mathbf{f}^\star := (f(\mathbf{x}_i) + e_i)_{i=1}^{s}$, where $\mathbf{e} := (e_i)_{i=1}^{s}$ is the noise vector. Then, the Discrete Fuzzy Transform of $\mathbf{f}$ and $\mathbf{f}^\star$ is given by

$$\mathbf{F}_n := D^{-1}K\mathbf{f}, \qquad \mathbf{F}_n^\star := D^{-1}K\mathbf{f}^\star,$$

respectively. From the previous relations, it follows that

$$\begin{aligned}
\|\mathbf{F}_n - \mathbf{F}_n^\star\|_2 &= \|D^{-1}K\mathbf{e}\|_2\\
&\leq \|D^{-1}\|_2\|K\|_2\|\mathbf{e}\|_2\\
&\leq \left[\min_{i=1,\dots,n}\left\{\sum_{j=1}^{s} A_i(\mathbf{x}_j)\right\}\right]^{-1} \sigma_{\max}(K)\|\mathbf{e}\|_2.
\end{aligned}$$

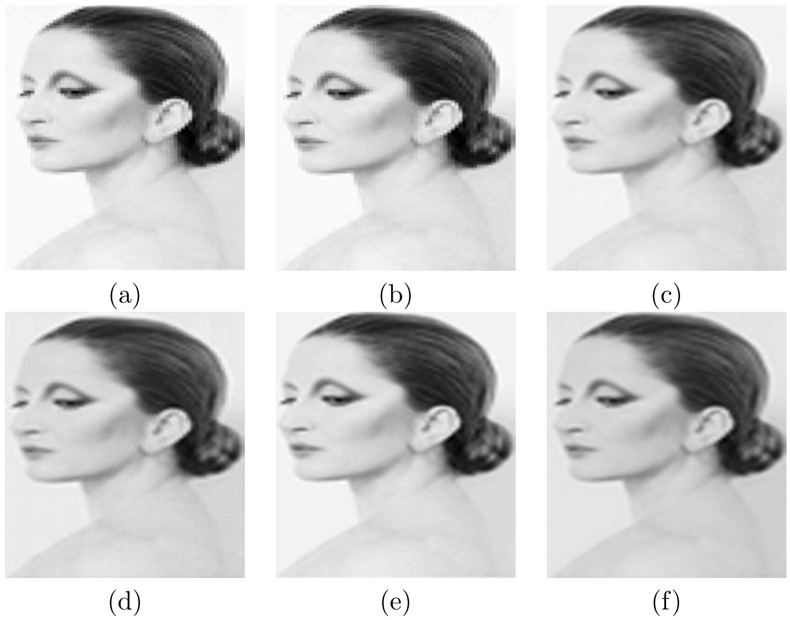

Figure 2: (a) Input 50×50 image and Inverse Fuzzy Transform (b) sampled on the same grid. Inverse Fuzzy Transform with (c,e) interpolating and (d,f) least-squares constraints on a (c,d) 200×200 and (e,f) 500×500 grid.

Then, the discrepancy between $\mathbf{F}_n$ and $\mathbf{F}_n^{\star}$ is proportional to the norm $\|\mathbf{e}\|_2$ of the perturbation vector and the maximum singular value of the matrix K.

3.2. Fuzzy Transform and least-squares approximation

Assuming $n < s$ and that the membership functions $\{A_i(\mathbf{x})\}_{i=1}^n$ are linearly independent, we search the map $\overline{f}(\mathbf{x}) := \sum_{i=1}^n \alpha_i A_i(\mathbf{x})$ that is the best least-squares approximation with respect to the f-values; i.e., the map $\overline{f} : \mathbb{R}^d \to \mathbb{R}$ that solves the minimization problem

$$\arg\min_{\overline{f}} \left\{ \sum_{j=1}^{s} |f(\mathbf{x}_j) - \overline{f}(\mathbf{x}_j)|^2 \right\}. \tag{6}$$

Inserting the expression of $\overline{f}$ in Equation (6) and deriving the quadratic functional with respect to the parameters $\alpha := (\alpha_i)_{i=1}^n$, the vector α is the unique solution to the *normal equation*

$$KK^T\alpha = K\mathbf{f} \qquad \longleftrightarrow \qquad \alpha = K^{\dagger}\mathbf{f}, \qquad K^{\dagger} := (KK^T)^{-1}K, \tag{7}$$

where K, $\mathbf{f}$ are the matrices introduced in Section 3.1 and $K^{\dagger}$ is the pseudo-inverse of K (Golub and VanLoan, 1989). Since K is the Gram matrix associated to a set of basis functions, it has full rank; in particular, KK^T is positive definite and (7) has a unique solution. If the number of sampled values and basis functions is the same (i.e., $n = s$), then the normal equation (7) becomes $K\alpha = \mathbf{f}$ and the function $\overline{f}$ interpolates the input f-values; i.e., $\overline{f}(\mathbf{x}_i) = f(\mathbf{x}_i)$, $i = 1, \ldots, s$.

The assumption that the membership functions are linearly independent is not restrictive; for instance, this condition is satisfied by the triangular, sinusoidal and Gaussian membership functions. If the membership functions are not linearly independent or their properties (e.g., number, regularity, support) are not good for the approximation scheme, then the previous framework is updated by selecting a set of basis functions that are different from the input membership maps. For instance, we can choose the functions $\{\varphi_i(\mathbf{x}) := \varphi(\|\mathbf{x} - \mathbf{x}_i\|_2)\}_{i=1}^n$ generated by a positive or semi-positive definite kernel $\varphi : \mathbb{R}^+ \to \mathbb{R}$ (Aronszajn, 1950; Schoelkopf and Smola, 2002; Wendland, 1995). For the membership functions that are generated by Gaussian or compactly-supported kernels, the coefficient matrix K is sparse and the memory allocation is linear in the number of samples.

Evaluation of the discrepancy error. Comparing the least-squares approximation with the Inverse Fuzzy Transform, we have that

$$
\begin{aligned}
|f_{F,n}(\mathbf{x}) - \overline{f}(\mathbf{x})| &= \left| \sum_{i=1}^n (F_i - \alpha_i) A_i(\mathbf{x}) \right| \\
&\leq \sum_{i=1}^n |F_i - \alpha_i| \, |A_i(\mathbf{x})| \\
&\leq \|\mathbf{F}_n - \alpha\|_1 \\
&\leq \sqrt{n} \|D^{-1}K\mathbf{f} - (K^TK)^{-1}K\mathbf{f}\|_2 \\
&\leq \sqrt{n} \| \left[D^{-1} - (KK^T) \right] K\|_2 |\mathbf{f}\|_2 \\
&\leq \sqrt{n} \|D^{-1} - (KK^T)^{-1}\|_2 \sigma_{\max}(K) \|\mathbf{f}\|_2 \\
&\leq \left[\left(\min_{i=1,\ldots,n} \{ \sum_{j=1}^s A_i(\mathbf{x}_j) \} \right)^{-1} + \lambda_{\min}^{-1}(KK^T) \right] \sigma_{\max}(K) \|\mathbf{f}\|_2 .
\end{aligned}
$$

For the interpolating case, the last error bound reduces to

$$
\begin{aligned}
|f_{F,n}(\mathbf{x}) - f(\mathbf{x})| &\leq \sqrt{n} \|\mathbf{F}_n - \alpha\|_2 \\
&= \|(D^{-1}K - K^{-1})\mathbf{f}\|_2 \\
&\leq \sqrt{n} \|D^{-1}K - K^{-1}\|_2 \|\mathbf{f}\|_2, \\
&\leq \sqrt{n} \left[\|D^{-1}\|_2 \|K\|_2 + \|K^{-1}\|_2 \right] \|\mathbf{f}\|_2 \\
&\leq \sqrt{n} \left[\frac{\lambda_{\max}(K)}{\min_{i=1,\ldots,n} \{ \sum_{j=1}^s A_i(\mathbf{x}_j) \}} + \frac{1}{\lambda_{\min}(K)} \right] \|\mathbf{f}\|_2 \qquad \mathbf{x} \in \mathbb{R}^d.
\end{aligned}
$$

It follows that the error $\|f_{F,n} - \overline{f}\|_\infty$ depends only on the input f-values and the spectral properties of the Gram matrix K.

(a) 50×50 (b) 100×100 (c) 260×260

Figure 3: (a) Input noisy image and (b-c) re-sampling of its Inverse Fuzzy Transform with least-squares constraints and at different resolutions.

Discrete Fuzzy Transform of $\overline{f}$ and linearity with respect to the approximation of $\mathcal{D}$. Once the approximation $\overline{f}:\mathbb{R}^d \rightarrow \mathbb{R}$ has been computed, we proceed in two different ways by (i) applying the continuous definition of the Fuzzy Transform to $\overline{f}$; (ii) computing the Discrete Fuzzy Transform of $\overline{f}$.

In the first case, we have that

$$\overline{F}_k := \frac{\int \overline{f}(\mathbf{x})A_k(\mathbf{x})d\mathbf{x}}{\int A_k(\mathbf{x})d\mathbf{x}} = \frac{\sum_{i=1}^{n} \alpha_i \int A_i(\mathbf{x})A_k(\mathbf{x})d\mathbf{x}}{\int A_k(\mathbf{x})d\mathbf{x}}, \qquad k = 1, \ldots, n.$$

Through this last relation, the evaluation of $\overline{F}_k$ requires to pre-compute (only once) the integral values and then calculate the component $\overline{F}_k$, $k = 1, \ldots, n$, for several $\overline{f}$. Furthermore, assuming that each membership function is generated by a kernel the previous computation reduces to evaluate the integrals of functions of one variable.

In the second case, we show that the Discrete Fuzzy Transform of the array $\overline{\mathbf{f}} := (\overline{f}(\mathbf{x}_i))_{i=1}^{s}$ is equal to the Discret Fuzzy Transform of $\mathbf{f} := (f(\mathbf{x}_i))_{i=1}^{s}$; i.e., the Discrete Fuzzy Transform is *invariant* to the interpolating and least-squares approximation of discrete data. To this end, we compute the Discrete Fuzzy Transform of $\overline{f}$ as

$$\overline{\mathbf{F}}_n := (\overline{F}_k)_{k=1}^{n}, \qquad \overline{F}_k := \frac{\sum_{i=1}^{s} \overline{f}(\mathbf{x}_i)A_k(\mathbf{x}_i)}{\sum_{i=1}^{s} A_k(\mathbf{x}_i)}, \qquad k = 1, \ldots, n.$$

If $\overline{f}$ interpolates the f values, then the Discrete Fuzzy Transform of f and $\overline{f}$ on $\mathcal{P}$ is the same; i.e., $\overline{\mathbf{F}}_n = \mathbf{F}_n$. If we consider the least-squares case, then we have that the k^{th} component is

$$\overline{F}_k = \frac{\sum_{i=1,\ldots,s}^{j=1,\ldots,n} \alpha_j A_j(\mathbf{x}_i)A_k(\mathbf{x}_i)}{\sum_{i=1}^{s} A_k(\mathbf{x}_i)}, \qquad k = 1, \ldots, n.$$

In matrix form, the previous relations can be written as $\overline{\mathbf{F}}_n = D^{-1}(KK^T)\alpha$, with $K := (A_k(\mathbf{x}_i))_{k=1,\ldots,n}^{i=1,\ldots,s} \in \mathbb{R}^{n \times s}$, and $D := \text{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n \times n}$ diagonal matrix with entries $d_k := \sum_{j=1}^{s} A_k(\mathbf{x}_j)$, $k = 1, \ldots, n$. Since $\alpha := (KK^T)^{-1}K\mathbf{f}$ (c.f., Equation (7)), we have that

$$\overline{\mathbf{F}}_n = D^{-1}(KK^T)(KK^T)^{-1}K\mathbf{f} = D^{-1}K\mathbf{f} =: \mathbf{F}_n.$$

From the last identity, it follows that the input map f and its least-squares approximation $\overline{f}$ have the same Discrete Fuzzy Transform.

To show that the computation of the Discrete Fuzzy Transform of $\overline{f}$ is still linear, it is enough to verify that the least-squares approximations is linear. To this end, let $\mathbf{f} := (f(\mathbf{x}_i))_{i=1}^{s}$, $\mathbf{g} := (g(\mathbf{x}_i))_{i=1}^{s}$ be the values of two maps $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ at the points of $\mathcal{P}$ and let $\overline{f}$, $\overline{g}$ be the corresponding least-squares approximations. Our aim is to prove that the least-squares approximation of the discrete set $\{af(\mathbf{x}_i) + bg(\mathbf{x}_i)\}_{i=1}^{s}$ is $a\overline{f} + b\overline{g}$ with $a, b \in \mathbb{R}$; i.e., $\overline{af + bg} = a\overline{f} + b\overline{g}$. From the previous assumptions, we have that

$$\begin{array}{lll} \overline{f}(\mathbf{x}) := \sum_{i=1}^{n} \alpha_i A_i(\mathbf{x}), & (KK^T)\alpha = K\mathbf{f}, & \alpha := (\alpha_i)_{i=1}^{n}, \\ \overline{g}(\mathbf{x}) := \sum_{i=1}^{n} \beta_i A_i(\mathbf{x}), & (KK^T)\beta = K\mathbf{g}, & \beta := (\beta_i)_{i=1}^{n}, \end{array}$$

and the linearity of the least-squares approximation follows from the equalities

$$(KK^T)(a\alpha + b\beta) = aK\mathbf{f} + bK\mathbf{g} = K(a\mathbf{f} + b\mathbf{g}).$$

An example of computation of the Inverse Fuzzy Transform and stability to noise is shown in Figure 2 and 3, respectively. In both cases, a different sampling set has been used to show the capability of recovering information outside the sampling grid $\mathcal{P}$.

Additivity of the Discrete Fuzzy Transform associated to $\mathcal{P}$. Since $\overline{f}$ is a continuous approximation of the f-values, we can compute the Discrete Fuzzy Transform of $\overline{f}$ with respect to a new set of points $\overline{\mathcal{P}} := \mathcal{P} \cup \{\mathbf{x}_{s+1}, \ldots, \mathbf{x}_q\} =: \mathcal{P} \cup \mathcal{P}^{\star}$. In this case, the difference between the Discrete Fuzzy Transform associated to $\mathcal{P}$ and $\overline{\mathcal{P}}$ is evaluated as

$$\begin{aligned} \overline{F}_k &:= \frac{\sum_{i=1}^{q} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i)}{\sum_{i=1}^{q} A_k(\mathbf{x}_i)} \\ &= \frac{\sum_{i=1}^{s} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i) + \sum_{i=s+1}^{q} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i)}{\sum_{i=1}^{s} A_k(\mathbf{x}_i) + \sum_{i=s+1}^{q} A_k(\mathbf{x}_i)} \\ &\leq \frac{\sum_{i=1}^{s} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i)}{\sum_{i=1}^{s} A_k(\mathbf{x}_i)} + \frac{\sum_{i=s+1}^{q} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i)}{\sum_{i=s+1}^{q} A_k(\mathbf{x}_i)} \\ &\leq^{(a)} F_k + \frac{\sum_{i=s+1}^{q} \overline{f}(\mathbf{x}_i) A_k(\mathbf{x}_i)}{\sum_{i=s+1}^{q} A_k(\mathbf{x}_i)} \\ &\leq^{(b)} F_k + \max_{i=s+1,\ldots,q} \{|\overline{f}(\mathbf{x}_i)|\}, \qquad k = 1, \ldots, n. \end{aligned}$$

From the inequality (a), it follows that $\overline{F}_k$ is bounded by F_k and the Discrete Fuzzy Transform associated to the set of new samples in $\mathcal{P}^{\star}$. From (b), the upper bound to $\overline{F}_k$ is given by F_k and the maximum of the absolute values of f at the points of $\mathcal{P}^{\star}$. Indeed, we estimate an upper bound to $\overline{\mathbf{F}}_n$ using the values of $\overline{\mathbf{F}}_n$ and the l_{∞}-norm of the new f-values $\{\overline{f}(\mathbf{x}_i)\}_{i=s+1}^{q}$.

3.3. Dual basis for the Inverse Fuzzy Transform

We rewrite the Inverse Fuzzy Transform as a linear combination of the f-values and the *dual basis* $\mathcal{B}^{FT} := \{B_i^{FT}(\mathbf{x})\}_{i=1}^{s}$, which is defined as the set of functions such that $f_{F,n}(\mathbf{x}) := \sum_{i=1}^{s} f(\mathbf{x}_i) B_i^{FT}(\mathbf{x})$. Introducing the function vector $\mathbf{A}(\mathbf{x}) := (A_i(\mathbf{x}))_{i=1}^{n} \in \mathbb{R}^{n\times 1}$, $\mathbf{x} \in \mathbb{R}^d$, we get that

$$f_{F,n}(\mathbf{x}) := \mathbf{F}_n^T \mathbf{A}(\mathbf{x}) = (D^{-1}K\mathbf{f})^T \mathbf{A}(\mathbf{x}) = \mathbf{f}^T \underbrace{K^T D^{-1} \mathbf{A}(\mathbf{x})}_{:=\mathbf{B}^{FT}(\mathbf{x})},$$

where the components of $\mathbf{B}^{FT}(\mathbf{x}) := K^T D^{-1}\mathbf{A}(\mathbf{x})$ are the *dual membership functions*

$$B_i^{FT}(\mathbf{x}) := \sum_{j=1}^{n} \frac{A_j(\mathbf{x}_i)}{\sum_{r=1}^{s} A_j(\mathbf{x}_r)} A_j(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^d, \qquad i = 1, \dots, s.$$

For the interpolating and least-squares case, the approximating function in Equation (6) and (7) can be written as

$$\overline{f}(\mathbf{x}) := \alpha^T \mathbf{A}(\mathbf{x}) = \left[(KK^T)^{-1}K^T\mathbf{f}\right]^T \mathbf{A}(\mathbf{x}) = \mathbf{f}^T \underbrace{K(KK^T)^{-1}\mathbf{A}(\mathbf{x})}_{:=\mathbf{B}^{LS}(\mathbf{x})}.$$

As done before, the components of $\mathbf{B}^{LS}(\mathbf{x})$ are the *dual basis functions* of the least-squares approximation $\overline{f}$. While the evaluation of $\mathbf{B}^{FT}(\mathbf{x})$ takes linear time in n, the computation of $\mathbf{B}^{LS}(\mathbf{x})$ requires to invert the positive-definite matrix KK^T. In most of the cases, each membership function $A_i(\mathbf{x})$, $i = 1, \dots, n$, decreases with the distance of its center $\mathbf{x}_i$ from the evaluation point $\mathbf{x} \in \mathbb{R}^d$. In this way, only the data in a neighborhood of $\mathbf{x}_i$ affects the evaluation of $A_i(\mathbf{x})$, thus providing a local approximation scheme. In a similar way, the dual basis functions are bell-shaped; in fact, $f_{F,n}$ and $\overline{f}$ are linear combinations of these maps with the f-values as coefficients. The bell-shape is also enforced by the term D^{-1} and $(KK^T)^{-1}$ in $f_{F,n}$ and $\overline{f}$, respectively. From the dual formulation, we deduce that the Inverse Fuzzy Transform preserves the local features of the input data.

4. Enriching the Discrete and Inverse Fuzzy Transform with geometric and confidence information

The invariance of the Discrete Fuzzy Transform with respect to the least-squares approximation suggests us to use this framework as basis for inserting additional information on the input data in the approximation scheme. In this way, the Fuzzy Transform is enriched with geometric information on $\mathcal{P}$ (Section 4.1) and confidence weights associated to the discrete f-values (Section 4.2). For instance, this is the case of the values of a function sampled on a manifold surface, whose dimension is lower than that of the embedding space; the pixels of an image; the points generated by physical processes; or the maps that are solutions of differential equations.

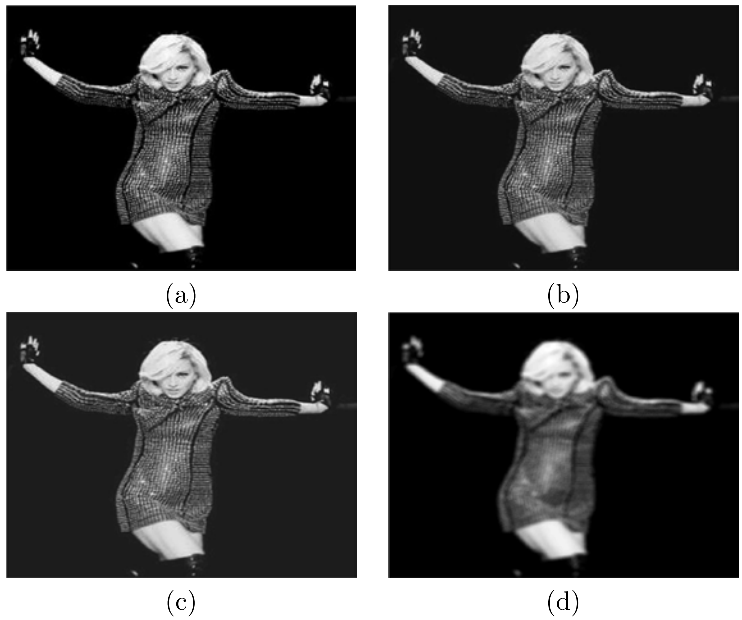

Figure 4: (a) Input image and (b-d) geometry-driven Inverse Fuzzy Transform with a different k-nearest neighbor graph used to compute the Laplacian matrix; i.e., (b) $k = 10$, (c) $k = 15$, and (d) $k = 20$.

4.1. Enriching the Fuzzy Transform with geometric information

Until now the approximation of the f-values has been driven by the only membership functions, thus neglecting the distribution of the input data $\mathcal{P}$. To introduce the geometric organization of $\mathcal{P}$, let us now define the approximation $\overline{f} : \mathbb{R}^d \to \mathbb{R}$ that minimizes two conflicting criteria: the *least-squares error* $\sum_{i=1}^{s} |f(\mathbf{x}_i) - \overline{f}(\mathbf{x}_i)|^2$ and the *intrinsic geometric term* $\|\overline{\mathbf{f}}\|_L^2 := \overline{\mathbf{f}}^T L \overline{\mathbf{f}}$, where L is a $s \times s$ semi-positive definite matrix and $\overline{\mathbf{f}} := (\overline{f}(\mathbf{x}_i))_{i=1}^{s}$ is the array of $\overline{f}$-values at the points of $\mathcal{P}$. Roughly speaking, the intrinsic penalty term must reflect the intrinsic geometry of the input data; as detailed at the end of this section, a possible solution is to choose L as the Laplacian matrix associated to the k-nearest neighbor graph of $\mathcal{P}$.

Assuming that the approximation $\overline{f}(\mathbf{x}) := \sum_{i=1}^{n} \alpha_i A_i(\mathbf{x})$ is a linear combination of the basis functions $\mathcal{A} := \{A_i(\mathbf{x})\}_{i=1}^{n}$, $\overline{f}$ is the solution to the minimization problem

$$\Phi(\alpha_1, \ldots, \alpha_n) := \sum_{i=1}^{s} |f(\mathbf{x}_i) - \overline{f}(\mathbf{x}_i)|^2 + \lambda \|\overline{\mathbf{f}}\|_L^2, \qquad \lambda \geq 0,$$

where λ is the trade-off between the least-squares and the intrinsic term. As λ increases, the effect of the geometry of the data increases and alters the complexity behavior of the dual basis functions; for $\lambda := 0$, we get the least-squares functional (6). Since $\overline{\mathbf{f}} := (\overline{f}(\mathbf{x}_i))_{i=1}^{s} := K^T \alpha$, where $K := (A_i(\mathbf{x}_j))_{i=1,\ldots,n}^{j=1,\ldots,s} \in \mathbb{R}^{n \times s}$

is the Gram matrix associated to the set $\mathcal{A}$, the functional Φ is rewritten as

$$\begin{aligned}\Phi(\alpha_1,\ldots,\alpha_n) &:= \|\mathbf{f}-\bar{\mathbf{f}}\|_2^2+\lambda\|\bar{\mathbf{f}}\|_L^2\\ &= \left[\mathbf{f}^T\mathbf{f}-2\alpha^T K\mathbf{f}+\alpha^T(KK^T)\alpha\right]+\lambda\|K^T\alpha\|_L^2\\ &= \alpha^T\left[K(I+\lambda L)K^T\right]\alpha-2\alpha^T K\mathbf{f}+\mathbf{f}^T\mathbf{f},\end{aligned}$$

whose normal equations $\nabla\Phi=\mathbf{0}$ are

$$K(I+\lambda L)K^T\alpha=K\mathbf{f}\quad\leftrightarrow\quad\alpha:=\left[K(I+\lambda L)K^T\right]^{-1}K\mathbf{f}. \tag{8}$$

If L is semi-positive definite, then $(I+\lambda L)$ is positive definite for any $\lambda\geq 0$. Since K has maximal rank (i.e., K is the Gram matrix associated to the membership functions), the matrix $K(I+\lambda L)K^T$ is positive definite for any $\lambda>0$ and α is the unique solution to Equation (8). Then, we derive the expression of the *geometry-driven Discrete Fuzzy Transform* $\mathbf{F}_n^G$ as follows

$$\begin{aligned}\mathbf{F}_n^G &:= D^{-1}(KK^T)\alpha\\ &= D^{-1}(KK^T)\underbrace{\left[K(I+\lambda L)K^T\right]^{-1}K\mathbf{f}}_{:=\mathbf{y}}.\end{aligned}$$

To compute $\mathbf{F}_n^G$ without explicitly inverting the matrix $K(I+\lambda L)K^T$, which is computationally unfeasible, the previous relation is rewritten as the system

$$\begin{cases}\left[K(I+\lambda L)K^T\right]\mathbf{y}=K\mathbf{f} & \text{(a)}\\ \mathbf{F}_n^G:=D^{-1}(KK^T)\mathbf{y} & \text{(b).}\end{cases} \tag{9}$$

Indeed, firstly we compute $\mathbf{y}$ by solving the sparse linear system (9a) in $O(s\log s)$ time; then, $\mathbf{F}_n^G$ is evaluated in linear time through (9b). Finally, the vector $\mathbf{F}_n^G:=(F_i^G)_{i=1}^n$ uniquely identifies the corresponding *geometry-driven Inverse Fuzzy Transform*, which is defined as $f_{F,n}^G(\mathbf{x}):=\sum_{i=1}^n F_i^G A_i(\mathbf{x})$. Figure 4 shows the level-sets of the geometry-driven Inverse Fuzzy Transform $f_{F,n}$ of the confidence map δ, defined on a 2D curve which is represented in (a). Since the highest confidence value is one, the reconstructed curve is achieved as $f_{F,n}^{-1}(1)$. Finally, notice that increasing k provides smoother results which might appear blurred (e.g., Figure 4(d)).

As suggested in (Sindhwani et al., 2006), a natural way to estimate the correlation among points and the geometry of the underlying data set is to use the graph Laplacian (Chung, 1997), which is defined as $L:=D-W$ where D is a diagonal matrix and W is the adjacency matrix induced by the k-nearest neighbor graph of $\mathcal{P}$. Here, the $s\times s$ matrices D and W are defined as

$$W_{ij}:=\begin{cases}\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|_2}{4h}\right), & \mathbf{x}_j\in\mathcal{N}_{\mathbf{x}_i},\\ 0 & \text{else,}\end{cases}\quad D_{ij}:=\begin{cases}\sum_{\mathbf{x}_j\in\mathcal{N}_{\mathbf{x}_i}}W_{ij}, & i=j,\\ 0 & \text{else,}\end{cases}$$

where h is the support of the Gaussian weight function and $\mathcal{N}_{\mathbf{x}_i}$ is the set of the k points of $\mathcal{P}$ that has the shortest Euclidean distance from $\mathbf{x}_i$, $i=1,\ldots,n$. Alternatively, we can use constant weights $W_{ij}:=1$ if (i,j) is an edge of the k-nearest

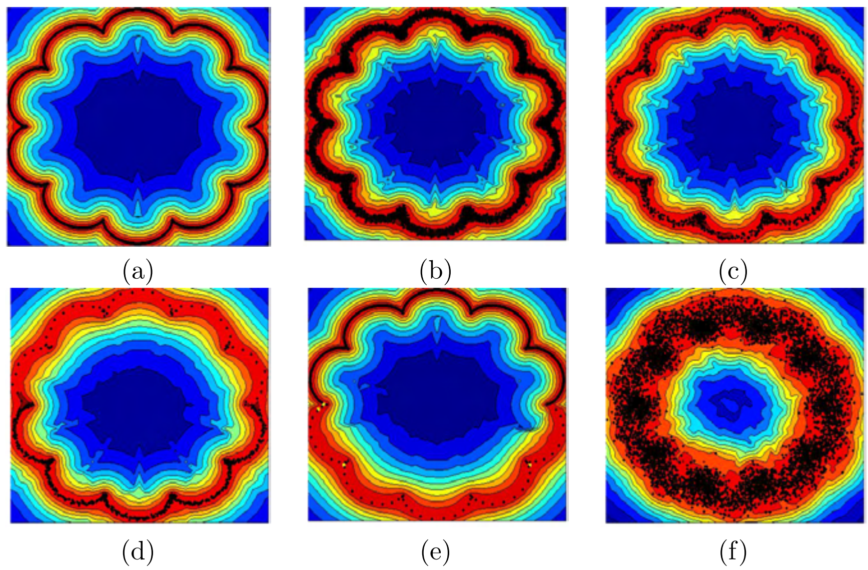

Figure 5: Level-sets of the confidence-driven Inverse Fuzzy Transform associated to a set of points (black dots), which is (a-e) characterized by a different sampling density and (f) affected by noise.

neighbor graph of $\mathcal{P}$. In both cases, the Laplacian matrix is sparse and the number of non-null element is $(k+1)\times n$. Finally, the k-nearest neighbor graph is efficiently computed in $O(kn\log n)$ time using efficient data structures (Arya et al., 1998).

4.2. Enriching the Fuzzy Transform with confidence information

In the following, we discuss how confidence maps associated to the input data can be used in the context of the Fuzzy Transform for data modeling and analysis. Let us assume that $\mathcal{P}$ is enriched with a *confidence map* $\delta : \mathcal{P} \rightarrow [0,1]$, which associates to each value $f(\mathbf{x}_i)$ or point $\mathbf{x}_i$ of $\mathcal{P}$ its degree of reliability $\delta(\mathbf{x}_i) := \delta_i$. This map is *a-priori* known, or provided by the acquisition process (e.g., through laser scanners), or computed by analyzing the data variability through a likelihood estimation (Jolliffe, 1986). A higher value of δ corresponds to a higher reliability. For more details on the computation of δ for point-sampled surfaces, we refer the reader to (Grigoryan and Rheingans, 2004; Pauly et al., 2004).

Assuming that δ_i is the confidence weight associated to the value $f(\mathbf{x}_i)$, we rewrite the least-squares functional (6) as

$$\Phi(\alpha_1,\dots,\alpha_n) := \sum_{i=1}^{s}\delta_i|f(\mathbf{x}_i)-\overline{f}(\mathbf{x}_i)|^2 = \sum_{i=1}^{s}\delta_i\left|f(\mathbf{x}_i)-\sum_{j=1}^{n}\alpha_j A_j(\mathbf{x}_i)\right|^2.$$

Then, the stationary points of Φ are the solutions to the linear equations

$$\sum_{i=1}^{s} \delta_i A_k(\mathbf{x}_i) f(\mathbf{x}_i) = \sum_{i=1,\ldots,s}^{j=1,\ldots,n} \delta_i A_j(\mathbf{x}_i) A_k(\mathbf{x}_i)\alpha_j, \qquad k = 1,\ldots,n.$$

These last relations can be written in matrix form as

$$(K\Delta K^T)\alpha = K\Delta\mathbf{f} \qquad \longleftrightarrow \qquad \alpha = (K\Delta K^T)^{-1}K\Delta\mathbf{f},$$

where $\Delta := \mathrm{diag}(\delta_1,\ldots,\delta_s) \in \mathbb{R}^{s\times s}$ is the diagonal matrix whose entries are the confidence values associated to the points of $\mathcal{P}$. Finally, the *confidence-driven Discrete Fuzzy Transform* is

$$\mathbf{F}_n^C := D^{-1}(KK^T)^{-1}\alpha = D^{-1}(KK^T)\underbrace{(K\Delta K^T)^{-1}K\Delta\mathbf{f}}_{:=\mathbf{y}}. \tag{10}$$

Assuming that $\delta_i \neq 0$, $i = 1,\ldots,s$, we have that Δ is invertible; from this property, it follows that $K\Delta K^T$ is positive definite and the solution α to (10) is unique. To compute $\mathbf{F}_n^C$ without explicitly inverting the matrix $K\Delta K^T$, the previous relation is rewritten as the system

$$\begin{cases} (K\Delta K^T)\mathbf{y} = K\Delta\mathbf{f} & \text{(a)} \\ \mathbf{F}_n^C := D^{-1}(KK^T)\mathbf{y} & \text{(b).} \end{cases} \tag{11}$$

To solve (11), firstly we compute $\mathbf{y}$ by solving the sparse linear system (11a) in $O(s\log s)$ time; then, $\mathbf{F}_n^C$ is evaluated in linear time through (11b). The values of the vector $\mathbf{F}_n^C := (F_i^C)_{i=1}^n \in \mathbb{R}^n$ uniquely identify the corresponding *confidence-driven Inverse Fuzzy Transform*, which is defined as $f_{F,n}^C(\mathbf{x}) := \sum_{i=1}^n F_i^C A_i(\mathbf{x})$. Finally, notice that we can combine the geometric- and confidence-driven Discrete Fuzzy Transform in a unique framework. Figure 5 shows the results of the confidence-driven Inverse Fuzzy Transform on a data set $\mathcal{P}$ with (a,b) a regular, (c,e) an irregular sampling density, and (f) noise. Until the noise remains low, the level sets of the corresponding confidence-driven Inverse Fuzzy Transform remain smooth and well-distributed around $\mathcal{P}$; with a high noise, the level-sets become almost circular and centered at the barycenter of $\mathcal{P}$.

5. Conclusions and future work

This paper has discussed the relation between the least-squares approximation and the Fuzzy Transform, its specialization to radial membership functions, and the invariance of the Discrete Fuzzy Transform to interpolating and least-squares constraints. Then, we have shown how geometric information on the input data and confidence weights associated to the f-values can be inserted in the computation of the Discrete Fuzzy Transform. The exploitation of this information, which is extracted with statistical and graph-based techniques, generally provides an Inverse Fuzzy Transform that recovers the features of the function underlying the input data. Furthermore, the main properties of the Fuzzy Transform are preserved by these enhanced formulations.

The proposed approach, which is oblivious of the dimension and organization of the input data, uses numerical linear algebra as main tool and requires the solution of sparse linear systems, thus resulting computationally feasible in real applications. Even though the $O(n \log n)$ computational cost of the geometry- and confidence-driven Discrete Fuzzy Transform is higher than the linear time needed to compute the Discrete Fuzzy Transform, the proposed extensions are meaningful in terms of cost and results.

As future work, we plan to analyze the analogies between the Fuzzy Transform and other approximation techniques, which include the Partition of the Unity and moving least-squares techniques, and to provide an interpretation of the proposed approach using residuated lattice or analogous paradigms.

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. In: Proc. of the ACM SIGMOD International Conference on Management of data. pp. 207–216.

Aronszajn, N., 1950. Theory of reproducing kernels. Transactions of the American Mathematical Society 68.

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., Wu, A. Y., 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. Journal of the ACM 45 (6), 891–923.

Chung, F. R. K., 1997. Spectral graph theory. American Mathematical Society.

Daubechies, I., 1992. Ten Lectures on Wavelets (CBMS - NSF Regional Conference Series in Applied Mathematics). Society for Industrial & Applied MathematicsFA.

Di Martino, F., Loia, V., Perfilieva, I., Sessa, S., 2008. An image coding/decoding method based on direct and inverse fuzzy transforms. International Journal of Approximate Reasoning 48 (1), 110–131.

Dyn, N., Levin, D., Rippa, S., 1986. Numerical procedures for surface fitting of scattered data by radial functions. SIAM Journal on Scientific and Statistical Computing 7(2), 639–659.

Farwig, R., 1986. Multivariate interpolation of arbitrarily spaced data by moving least squares methods. Journal of Computational and Applied Mathematics 16 (1), 79–93.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. Artifical Intelligence Magazine 17 (3), 37–54.

Golub, G., VanLoan, G., 1989. Matrix Computations. John Hopkins University Press, 2nd Edition.

Grigoryan, G., Rheingans, P., 2004. Point-based probabilistic surfaces to show surface uncertainty. IEEE Transactions on Visualization and Computer Graphics 10 (5), 564–573.

Hong, T.-P., Lin, K.-Y., Wang, S.-L., 2003. Fuzzy data mining for interesting generalized association rules. Fuzzy Sets and System 138 (2), 255–269.

Jolliffe, I. T., 1986. Principal component analysis. In: Principal Component Analysis. Springer Verlag.

Micchelli, C. A., 1986. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. Constructive Approximation 2, 11–22.

Mitra, S., Pal, S., Mitra, P., 2002. Data mining in soft computing framework: a survey. IEEE Transactions on Neural Networks 13 (1), 3–14.

Novák, V., Perfilieva, I., Dvořák, A., Chen, G., Wei, Q., Yan, P., 2008. Mining pure linguistic associations from numerical data. Internationals Journal of Approximate Reasoning 48 (1), 4–22.

Pauly, M., Mitra, N., Guibas, L., 2004. Uncertainty and variability in point cloud surface data. Proc. of the Symposium on Point-based Graphics, 77–84.

Perfilieva, I., 2005. Fuzzy transforms and their applications to data compression. In: International Conference on Fuzzy Systems. pp. 294–299.

Perfilieva, I., 2006. Fuzzy transforms: Theory and applications. Fuzzy Sets and Systems 157 (8), 993 – 1023.

Perfilieva, I., Novák, V., Dvořák, A., 2008. Fuzzy transform in the analysis of data. International Journal of Approximate Reasoning 48 (1), 36–46.

Piatetsky-Shapiro, G., 2000. Knowledge discovery in databases: Ten years after. SIGKDD Explorations 1 (2), 59–61.

Schoelkopf, B., Smola, A. J., 2002. Learning with Kernels. The MIT Press.

Sindhwani, V., Belkin, M., Niyogi, P., 2006. Geometric basis of semi-supervised learning. In: Semi-Supervised Learning. pp. 209–226.

Stepnicka, M., Polakovic, O., 2009. A neural network approach to the fuzzy transform. Fuzzy Sets and Systems 160 (8), 1037–1047.

Wendland, H., 1995. Real piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Advances in Computational Mathematics 4(4), 389–396.

Zhang, L., Shi, Y., Yang, X., 2006. Association-rule knowledge discovery by using a fuzzy mining approach. International Journal of Business Intelligence and Data Mining 1 (4), 417–429.