

# Dimensionality Reduction and Its F-Transform Representation

Jiří Janeček<sup>a</sup> and Irina Perfiljeva<sup>b</sup>

<sup>a</sup>Department of Mathematics, Faculty of Science, University of Ostrava,  
30. dubna 22, 701 03 Ostrava 1, Czech Republic, jiri.janecek@osu.cz

<sup>b</sup>Institute for Research and Applications of Fuzzy Modeling, University of Ostrava,  
30. dubna 22, 701 03 Ostrava 1, Czech Republic, irina.perfiljeva@osu.cz

## Abstract

The aim of this contribution is to connect two techniques: the F-transform on a space with a fuzzy partition and dimensionality reduction. We show how a space with a fuzzy partition can be characterized in terms of the graph Laplacian. After making the spectral analysis of the corresponding Laplacian matrix we compute its eigenvectors and show their relevance to the F-transform basic functions.

**Keywords:** F-transform, Graph Laplacian, Dimensionality reduction.

## 1 Introduction

It is a very challenging task to analyze and process “big data”. In particular, a problem is how to work with a dataset with more than say, a thousand features? Having a high number of variables is both a godsend and a curse. It might take days or months to perform any meaningful analysis with reasonable computational power. There should exist a better way to deal with high dimensional data for quick pattern extraction and further insights.

It seems that the only solution is using dimensionality reduction techniques. This concept helps to reduce the number of features without having to lose much information. Below are listed some benefits of applying dimensionality reduction to a dataset:

- Space required to store the data is reduced,
- Less dimensions lead to less computation/training time,
- Performance of algorithms is better,
- It helps in visualizing data.

Dimensionality reduction per se can be done in two different ways: by keeping the most relevant variables from the original dataset (*feature selection*); by finding a smaller set of new variables, containing basically the same information as the input variables (*dimensionality reduction*).

In this contribution we follow the second way and propose a new technique for the dimensionality reduction based on the F(fuzzy)-transforms. In this respect, the reduction is performed by dividing data into uniformly distributed observable pieces (fuzzy partition) and then, projecting each piece on eigenvectors of the corresponding Laplacian. The latter indicates importance of this particular region with respect to the whole dataset.

We use [1] for a general description of a dimensionality reduction technique, that is based on the graph Laplacian. We stem from [2, 3] for the detailed description of a higher order F-transform. Finally, we are based on [4, 5] where initial steps towards the F-transform based dimensionality reduction were attempted.

In this contribution, we give necessary technical details of the proposed approach. By this we mean a description of a fuzzy partition of data, establishing a characterization of a uniform fuzzy partition in terms of closeness matrix, computation of the corresponding Laplacian matrices and characterization of their eigenvalues and eigenvectors. The latter are used for specifying principal F-transform components. The reduced representation is then comprised of F-transform components.

The structure of this contribution is as follows. In Section 2, we give necessary details of dimensionality reduction based on a graph Laplacian. In Section 3, we remind the basics of the F-transform. In Section 4, we show a connection between a fuzzy partition and the dimensionality reduction.

## 2 Dimensionality Reduction Formally

We start with a formal presentation of the problem of dimensionality reduction as it appeared in [1]:

For a given set of points  $\vec{x}_1, \dots, \vec{x}_q$  from the space  $\mathbb{R}^l$ , find a set of corresponding representatives (images)  $\vec{y}_1, \dots, \vec{y}_q$  from the space  $\mathbb{R}^m$ , where  $m \ll l$ , such that the following cost function (case  $m = 1$ ) is minimized:

$$\sum_{i,j=1}^q (y_i - y_j)^2 w_{ij}, \quad (1)$$

where  $w_{ij}$  is a measure of closeness between the prototypes  $\vec{x}_i$  and  $\vec{x}_j$ .

The meaning of the cost function consists in preserving local similarities among the prototypes.

### 2.1 Graph-based approach to the assignment of weights

In this section, we give necessary details regarding the assignment of weights as measures of closeness. We use the graph-based semantics following [1].

Let  $G = (V, E)$  be a weighted graph where each single vertex corresponds to a data point from  $\{\vec{x}_1, \dots, \vec{x}_q \in \mathbb{R}^l\}$ , so that  $|V| = q$ . The edges are assigned weights according to a certain measure of closeness between the vertices connected by them. In [1], the two options for weighting the connections between vertices are considered:

- a) *Heat kernel*. If vertices  $i$  and  $j$  are connected, then

$$w_{ij} = e^{-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{t}},$$

otherwise  $w_{ij} = 0$ . Parameter  $t \in \mathbb{R}$  should be specified beforehand.

- b) *Simple-minded assignment*.  $w_{ij} = 1$ , if vertices  $i$  and  $j$  are connected by an edge, and  $w_{ij} = 0$ , otherwise.

We assume that graph  $G$  is connected, otherwise we treat each component in the same manner separately.

### 2.2 Dimensionality reduction algorithm

For simplicity, we consider the reduction mapping (embedding) onto  $\mathbb{R}^1$ , i.e., the case  $m = 1$ . The algorithm [1, 4] runs as follows:

*Step 1.* Compute the *generalized* eigenvalues and eigenvectors for the Laplacian matrix  $L$  in accordance with

$$L\vec{f} = \lambda D\vec{f}, \quad (2)$$

where  $D$  is the diagonal matrix such that

$$d_{ii} = \sum_{j=1}^q w_{ij},$$

and  $L = D - W$ .

*Step 2.* If  $\vec{f}_0, \dots, \vec{f}_{q-1}$  are the solutions of equation (2), ordered according to their eigenvalues, then, we leave out the eigenvector  $\vec{f}_0$ , corresponding to the eigenvalue 0, and use the eigenvector  $\vec{f}_1$  corresponding to the second smallest eigenvalue. The solution (embedding in 1-dimensional Euclidean space) is given by

$$\vec{x}_i \mapsto \vec{f}_1(i), \quad i = 1, \dots, q. \quad (3)$$

We remark that by preserving local information in the embedding, the algorithm naturally emphasizes the data clusters. Below, we propose a justification to the above given algorithm.

Because the matrix  $W$  is symmetric, the cost function in (1) can be rewritten as follows [1]:

$$\frac{1}{2} \sum_{i,j=1}^q (y_i - y_j)^2 w_{ij} = \vec{y}^\top L \vec{y}. \quad (4)$$

Hence,  $L$  is symmetric and positive semidefinite matrix. It is known that vector  $\vec{y}$  that minimizes  $\vec{y}^\top L \vec{y}$  corresponds to the minimal eigenvalue solution to the generalized eigenvalue problem (2), i.e.,

$$L\vec{y} = \lambda D\vec{y}.$$

Therefore,

$$\min_{\vec{y}(\lambda)} \vec{y}^\top L \vec{y} = \min_{\vec{y}(\lambda)} \lambda \vec{y}^\top D \vec{y}.$$

Because any eigenvector is determined up to a nonzero multiplier, we can add the normalization requirement  $\vec{y}^\top D \vec{y} = 1$  to (2), so that the considered problem reduces to

$$\min_{\vec{y}^\top D \vec{y} = 1} \vec{y}^\top L \vec{y},$$

and the normalized solution is determined by the minimal generalized eigenvalue.

By (4), this minimum eigenvalue solution minimizes the cost function in (1). By [1], the minimal eigenvalue of the generalized eigenvalue problem (2) is equal to 0, and the constant vector  $\vec{y} = \vec{1}$  is the only (up to a multiplier) eigenvector for  $\lambda = 0$  (provided that the graph  $G$  is connected). Therefore, in *Step 2*, we select as a non-trivial solution the eigenvector  $\vec{f}_1$ , corresponding to the second smallest eigenvalue.

### 3 Fuzzy Transform

The substantial characterization of the technique of F-transform (originally, *fuzzy transform*, [2]) is that it is an integral transform that uses a *fuzzy partition* of a universe of discourse (usually, a bounded interval of reals  $[a, b]$ ). The F-transform has two phases: direct and inverse. The direct F-transform (FzT) is applied to functions from  $L^2([a, b])$  and maps them linearly onto (finite) sequences of numeric/functional components. Each component is a weighted orthogonal projection of a given function on a certain linear subspace of  $L^2([a, b])$ . Dimensionality reduction by the F-transform is thought as a projection of original data onto the set of the F-transform components, see [2] for the details.

Below, we recall some definitions from [2, 3].

#### 3.1 Fuzzy Partition

Let  $[a, b]$  be an interval on the real line  $\mathbb{R}$ . Fuzzy sets on  $[a, b]$  are identified by their membership functions; i.e., they are mappings from  $[a, b]$  into  $[0, 1]$ .

**Definition 1.** Let  $[a, b]$  be an interval on  $\mathbb{R}$ ,  $n \geq 2$ , and let  $x_0, x_1, \dots, x_n, x_{n+1}$  be nodes such that  $a = x_0 \leq x_1 < \dots < x_n \leq x_{n+1} = b$ . We say that fuzzy sets  $A_1, \dots, A_n : [a, b] \rightarrow [0, 1]$ , which are identified with their membership functions, constitute a fuzzy partition of  $[a, b]$  if for  $k = 1, \dots, n$ , they fulfill the following conditions:

1. (locality) -  $A_k(x) = 0$  if  $x \in [a, x_{k-1}] \cup [x_{k+1}, b]$ ,
2. (continuity) -  $A_k(x)$  is continuous,
3. (covering) -  $A_k(x) > 0$  if  $x \in (x_{k-1}, x_{k+1})$

The membership functions  $A_1, \dots, A_n$  are called *basic functions*.

In the case of discrete data  $P = \{p_1, \dots, p_l\} \subset [a, b]$ , we call the set  $P$  *sufficiently dense* with respect to the given partition, if

$$\forall k \quad \exists j: A_k(p_j) > 0.$$

This means that each partition unit covers at least one point.

#### 3.2 Hilbert Space with Weighted Inner Product

Let us fix  $[a, b]$  and its fuzzy partition  $A_1, \dots, A_n$ , with nodes  $x_0, x_1, \dots, x_n, x_{n+1}$ ,  $n \geq 2$ . Let  $k$  be a fixed integer from  $\{1, \dots, n\}$ , and let  $L_2(A_k)$  ( $L_2([a, b])$ ) be a set

of square-integrable functions on  $[x_{k-1}, x_{k+1}]$  ( $[a, b]$ ) with inner products

$$\begin{aligned} \langle f, g \rangle_k &= \int_{x_{k-1}}^{x_{k+1}} f(x)g(x)A_k(x)dx, \\ \langle f, g \rangle &= \int_a^b f(x)g(x)dx. \end{aligned}$$

Spaces  $L_2(A_k)$ ,  $k = 1, \dots, n$ , and  $L_2([a, b])$  are Hilbert spaces on the corresponding domains. The functions  $f, g \in L_2(A_k)$  are *orthogonal* in  $L_2(A_k)$ , if  $\langle f, g \rangle_k = 0$ . The function  $f \in L_2(A_k)$  is orthogonal to a subspace  $B$  of  $L_2(A_k)$  if  $\langle f, g \rangle_k = 0$  for all  $g \in B$ .

Let us denote by  $L_2^m(A_k)$  a linear subspace of  $L_2(A_k)$  with the basis given by orthogonal polynomials  $P_k^0, P_k^1, P_k^2, \dots, P_k^m$ . For simplicity, we consider

$$P_k^0 = 1|_{[x_{k-1}, x_{k+1}]}, \quad P_k^1 = x - x_k|_{[x_{k-1}, x_{k+1}]},$$

according to [3]. The orthogonality of the considered basis polynomials gives

$$\langle P_k^0, P_k^1 \rangle = \int_{x_{k-1}}^{x_{k+1}} (x - x_k)A_k(x)dx = 0. \quad (5)$$

The following lemma gives analytic representation of the orthogonal projection on the subspace  $L_2^m(A_k)$ .

**Lemma 1** ([3]). Let function  $F_k^m$  be the orthogonal projection of  $f \in L_2(A_k)$  on  $L_2^m(A_k)$ . Then,

$$F_k^m = c_{k,0}P_k^0 + c_{k,1}P_k^1 + \dots + c_{k,m}P_k^m, \quad (6)$$

where for all  $i = 0, 1, \dots, m$ ,

$$c_{k,i} = \frac{\langle f, P_k^i \rangle_k}{\langle P_k^i, P_k^i \rangle_k} = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)P_k^i(x)A_k(x)dx}{\int_{x_{k-1}}^{x_{k+1}} P_k^i(x)P_k^i(x)A_k(x)dx}. \quad (7)$$

The  $n$ -tuple  $(F_1^m, \dots, F_n^m)$  is the  $F^m$ -transform of  $f$  with respect to  $A_1, \dots, A_n$ , or formally,

$$F^m[f] = (F_1^m, \dots, F_n^m).$$

$F_k^m$  is called the  $k^{th}$   $F^m$ -transform component of  $f$ .

### 4 Dimensionality Reduction Determined by a Fuzzy Partition

In this section, we show that the direct  $F$ -transform fully agrees with the dimensionality reduction strategy. In more detail, we show that discrete (vector) versions of the basis functions in  $L_2^m(A_k)$  correspond to eigenvectors of the proper selected Laplacian matrix. Therefore, the  $F^m$ -transform components are projections on these eigenvectors.

At first, we specify a dataset that will be characterized by the  $F$ -transform-based low-dimensional representation in the form of embedding maps and finally, by components. This dataset will be connected with a discrete representation of a function, say  $f$  on some domain (subset of the Euclidean  $\mathbb{R}^l$ ). For simplicity, we assume that the domain is an interval  $[a, b]$  of the real line, and the function  $f$  is given on a discrete set, say  $P$  of points where  $P \subseteq [a, b]$ . Then, we assume that  $[a, b]$  is partitioned into the collection of fuzzy sets  $A_1, \dots, A_n$ , and the set  $P$  is sufficiently dense with respect to this partition.

Moreover, we assume that for every  $k$ ,  $1 \leq k \leq n$ , there is one point  $x_k \in P$  (we call it *node*) such that  $A_k(x_k) = 1$  and  $A_j(x_k) = 0$ ,  $j \neq k$ . Further on, we distinguish between points (denoted by using letter  $p$ ) and nodes (denoted by using letter  $x$ ).

#### 4.1 Matrix representation of a fuzzy partition

The chosen partition determines a geometry of the set  $P$ , given by the following directed weighted graph  $D = (V(D), A(D))$ . Each vertex from  $V(D)$  corresponds to one point in  $P$ , and if a point  $p_i \in P$  is covered by basic function  $A_k$ , i.e.  $A_k(p_i) > 0$ , then the vertex  $i$  corresponding to  $p_i$  is connected by the directed edge  $(k, i)$  with the vertex  $k$ , corresponding to the node  $x_k$ . Let us emphasize that there is no connection between points, if there is no node among them. The weight of the directed edge  $(k, i)$  is equal to  $A_k(p_i)$ . It is easy to see that graph  $D$  is covered by  $n$  weakly connected components  $D_1, \dots, D_n$ , each has a “star” shape. In more detail,  $D_k = (V(D_k), A(D_k))$  is a subgraph of  $D$  associated with a “central” vertex (node)  $k$ . Then, the vertex set  $V(D_k)$  contains vertex  $k$  and moreover, vertex  $i$  belongs to  $V(D_k)$ , if  $(k, i) \in A(D_k)$ . Therefore, the set of directed edges  $A(D_k)$  consists of the edges  $(k, i)$  where  $i \in V(D_k)$  containing also the loop  $(k, k)$ . We will continue with each connected component  $D_k$  separately, and construct the low-dimensional representation of the part of the dataset corresponding to it (covered by the  $A_k$ ).

Let a finite set of points  $\{p_1, \dots, p_{l_k}\}$  be covered by  $A_k$  (a fuzzy partition unit). Let us remark that the node  $x_k$  is among these points. Let  $f_i = f(p_i)$ ,  $1 \leq i \leq l_k$ . The set  $X_k$  of data points allocated for the low-dimensional representation is  $\{(p_1, f_1), \dots, (p_{l_k}, f_{l_k})\}$ . Let  $D_k$  be the corresponding directed subgraph of  $D$  and  $G_k = (V(G_k), E(G_k))$  be the corresponding ordinary weighted graph derived from  $D_k$  after removing edge directions. Therefore,  $V(G_k) = V(D_k)$ , and if  $(k, i) \in A(D_k)$ , then  $\{k, i\} \in E(G_k)$ . In the weight matrix  $W$  corresponding to  $G_k$ , each edge, connecting  $i$  and  $k$ , is represented by the two symmetrical ele-

ments  $w_{ik} = w_{ki} = A_k(p_i)$ , the loop is represented by one diagonal entry  $w_{kk} = A_k(x_k) = 1$ . Therefore,  $W$  is a symmetrical  $l_k \times l_k$  matrix:

$$W = \begin{bmatrix} & & & a_1 & & & \\ & & & a_2 & & & \\ & & & \vdots & & & \\ & & & a_{k-1} & & & \\ a_1 & \dots & a_{k-1} & 1 & a_{k+1} & \dots & a_{l_k} \\ & & & a_{k+1} & & & \\ & & & \vdots & & & \\ & & & a_{l_k} & & & \end{bmatrix},$$

where  $a_i = A_k(p_i)$ . In the above given representation, the matrix blanks are considered as zeros.

The corresponding Laplacian matrix  $L = D - W$  has the form:

$$L = \begin{bmatrix} a_1 & & & -a_1 & & & \\ & \ddots & & \vdots & & & \\ & & a_{k-1} & -a_{k-1} & & & \\ -a_1 & \dots & -a_{k-1} & s_k - 1 & -a_{k+1} & \dots & -a_{l_k} \\ & & -a_{k+1} & a_{k+1} & & & \\ & & \vdots & & & \ddots & \\ & & -a_{l_k} & & & & a_{l_k} \end{bmatrix},$$

where  $D$  is a diagonal matrix such that

$$d_{ii} = \sum_{j=1}^k w_{ij},$$

and

$$D = \begin{bmatrix} a_1 & & & & & & \\ & \ddots & & & & & \\ & & a_{k-1} & & & & \\ & & \sum_{i=1}^{l_k} a_i & & & & \\ & & & a_{k+1} & & & \\ & & & & \ddots & & \\ & & & & & & a_{l_k} \end{bmatrix}.$$

**Theorem 1.** Let the component  $G_k$  be fixed and the corresponding matrices  $W, D, L$  are as above. Then, the generalized spectrum of Laplacian matrix  $L$  consists of 0, 1 and  $2 - \frac{1}{s_k}$ , where  $s_k = \sum_{i=1}^{l_k} A_k(p_i)$ . The multiplicity of the spectral eigenvalue 1 is  $l_k - 2$ .

*Proof.* The adjacency matrix  $W$  has the form written above. To analyze the spectrum of the Laplacian matrix  $L$ , used in the algorithm, we need to transform

the generalized eigenvalue problem

$$L\vec{y} = \lambda D\vec{y}$$

into a standard one in the simplest way possible – starting with the equality

$$L = D - W,$$

to avoid computations with the matrix  $L$ :

$$L\vec{y} = D\vec{y} - W\vec{y} = \lambda D\vec{y},$$

that gives

$$W\vec{y} = (1 - \lambda)D\vec{y}.$$

After the multiplication by the matrix  $D^{-1}$  from the left, we get

$$D^{-1}W\vec{y} = (1 - \lambda)\vec{y}.$$

Substituting

$$\mu = 1 - \lambda,$$

we obtain the standard eigenvalue problem

$$D^{-1}W\vec{y} = \mu\vec{y},$$

that is

$$(D^{-1}W - \mu I)\vec{y} = \vec{0},$$

requiring that

$$\det(D^{-1}W - \mu I) = 0,$$

where  $I$  is the identity matrix and  $\vec{y} \neq \vec{0}$  is the eigenvector of the matrix  $D^{-1}W$  corresponding to the eigenvalue  $\mu$ .

Let  $A_{0,\dots,h;0,\dots,h}$  denote the minor of the matrix  $D^{-1}W - \mu I$  (the determinant of its square submatrix such that the columns and rows  $0, \dots, h$  are deleted), then, expanding the determinant along the first rows, for all  $h = 0, \dots, k-2$ , it holds:

$$A_{0,\dots,h;0,\dots,h} = -\mu \left( A_{0,\dots,h+1;0,\dots,h+1} - \frac{a_{h+1}}{s_k} (-\mu)^{k-h-2} \right),$$

and

$$A_{0,\dots,k-1;0,\dots,k-1} = (-\mu)^{l_k-k-1} \cdot \left[ \left( \frac{1}{s_k} - \mu \right) (-\mu) - \sum_{j=k+1}^{l_k} \frac{a_j}{s_k} \right],$$

where

$$a_j = A_k(p_j).$$

Then we can see that  $\det(D^{-1}W - \mu I)$  can be rewritten in the form

$$(-\mu)^{l_k-2} \cdot R'(\mu), \quad (8)$$

which implies that there are  $l_k-2$  eigenvalues (denoted as  $\mu_2$ ) equal to 0 and that  $R'(\mu)$  has two more algebraic roots to be determined.

For simplicity, let us suppose that the vertex with the index 1 was assigned to the node  $x_k$  (the graph remains the same but we regulate the index assignment). Then  $k = 1$  and

$$\begin{aligned} \det(D^{-1}W - \mu I) &= A_{0,\dots,0;0,\dots,0} = \\ &= (-\mu)^{l_k-2} \cdot \left[ \left( \frac{1}{s_k} - \mu \right) (-\mu) - \sum_{j=2}^{l_k} \frac{a_j}{s_k} \right] = \\ &= (-\mu)^{l_k-2} \cdot R'(\mu), \end{aligned}$$

implying

$$\begin{aligned} R'(\mu) &= \left( \frac{1}{s_k} - \mu \right) (-\mu) - \sum_{j=2}^{l_k} \frac{a_j}{s_k} = \\ &= \mu^2 - \frac{\mu}{s_k} - \sum_{j=2}^{l_k} \frac{a_j}{s_k} = 0, \end{aligned}$$

being a quadratic equation with two roots:  $\mu_1 = 1$  and  $\mu_3 = \frac{1}{s_k} - 1$ . Recall that  $\mu_2 = 0$ . Now we know all the eigenvalues of the matrix  $D^{-1}W$ .

Substituting  $\lambda = 1 - \mu$ , we get all generalized eigenvalues of the Laplacian matrix  $L$ :  $\lambda_1 = 0$ ,  $\lambda_2 = 1$  and  $\lambda_3 = 2 - \frac{1}{s_k}$ . In the meaningful case  $l_k > 2$ ,  $s_k > 1$  and  $\lambda_3 \in (1, 2)$ , hence  $\lambda_2 < \lambda_3$ .  $\square$

## 4.2 Eigenvectors of Laplacian matrix

The eigenvector  $\vec{y}_1$ , corresponding to the generalized eigenvalue  $\lambda_1 = 0$ , is trivially equal to  $\vec{1}$ . Therefore, we will be focused on the eigenvectors corresponding to the generalized eigenvalues  $\lambda_2 = 1$  and  $\lambda_3 = 2 - \frac{1}{s_k}$ .

**Lemma 2.** *For the fixed component  $G_k$ , the eigenspace, corresponding to the second smallest generalized eigenvalue  $\lambda_2 = 1$ , is as follows:*

$$S_2 = \left\{ \vec{y} \in \mathbb{R}^{l_k} \mid \sum_{i=1}^{l_k} a_i y_i = 0 \quad \& \quad y_k = 0 \right\}. \quad (9)$$

*Proof.* It holds:

$$L\vec{y} = 1 \cdot D\vec{y},$$

that is

$$(D - W)\vec{y} = D\vec{y},$$

which gives

$$W\vec{y} = \vec{0}.$$

It means that

$$\begin{bmatrix} a_1 y_k \\ a_2 y_k \\ \vdots \\ a_{k-1} y_k \\ \sum_{i=1}^{l_k} a_i y_i \\ a_{k+1} y_k \\ \vdots \\ a_{l_k} y_k \end{bmatrix} = \vec{0},$$

where  $a_i = A_k(p_i) > 0$  and  $a_k = 1$ . From that, we deduce that

$$\forall i = 1, \dots, k-1, k+1, \dots, l_k : a_i y_k = 0,$$

and hence

$$y_k = 0.$$

From the  $k$ -th row we deduce that

$$\begin{aligned} \sum_{\substack{i=1 \\ i \neq k}}^{l_k} a_i y_i &= \sum_{\substack{i=1 \\ i \neq k}}^{l_k} a_i y_i + 1 \cdot 0 = \sum_{\substack{i=1 \\ i \neq k}}^{l_k} a_i y_i + a_k y_k = \\ &= \sum_{i=1}^{l_k} a_i y_i = 0. \end{aligned}$$

The last equality describes a hyperplane in the space  $\mathbb{R}^{l_k}$ , that is a subspace of the dimension  $l_k - 1$ , containing the origin. But the constraint  $y_k = 0$  reduces the dimension of that space to  $l_k - 2$  (there is only the trivial intersection with the  $k$ -th coordinate axis), because the algebraic multiplicity of  $\lambda$  is  $l_k - 2$  and due to symmetry of the matrix  $L$ , it equals to its geometric multiplicity as well.

To summarize: the eigenspace of the generalized eigenvalue 1 is characterized as follows:

$$S_2 = \left\{ \vec{y} \in \mathbb{R}^{l_k} \left| \sum_{i=1}^{l_k} a_i y_i = 0 \quad \& \quad y_k = 0 \right. \right\}.$$

□

**Lemma 3.** For the fixed component  $G_k$ , the eigenvector, corresponding to the largest generalized eigenvalue  $\lambda_3 = 2 - \frac{1}{s_k}$ , has the form

$$\vec{y}_3 = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \frac{1}{s_k} - 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (10)$$

*Proof.* It holds:

$$L\vec{y} = \left(2 - \frac{1}{s_k}\right) D\vec{y},$$

that is

$$D\vec{y} - W\vec{y} = \left(2 - \frac{1}{s_k}\right) D\vec{y},$$

which gives

$$W\vec{y} = \left(\frac{1}{s_k} - 1\right) D\vec{y},$$

that can be, after multiplication by  $D^{-1}$  from the left, rearranged into the form:

$$\left(D^{-1}W - \left(\frac{1}{s_k} - 1\right)I\right)\vec{y} = \vec{0},$$

where  $I$  is the identity matrix. The matrix  $D^{-1}W - \left(\frac{1}{s_k} - 1\right)I$  has the form:

$$\begin{bmatrix} 1 - \frac{1}{s_k} & & & 1 & & & \\ & \ddots & & \vdots & & & \\ & & 1 - \frac{1}{s_k} & 1 & & & \\ \frac{a_1}{s_k} & \dots & \frac{a_{k-1}}{s_k} & 1 & \frac{a_{k+1}}{s_k} & \dots & \frac{a_{l_k}}{s_k} \\ & & & 1 & 1 - \frac{1}{s_k} & & \\ & & & \vdots & & \ddots & \\ & & & 1 & & & 1 - \frac{1}{s_k} \end{bmatrix},$$

which gives

$$\begin{bmatrix} \left(1 - \frac{1}{s_k}\right)y_1 + y_k \\ \vdots \\ \left(1 - \frac{1}{s_k}\right)y_{k-1} + y_k \\ \sum_{\substack{i=1 \\ i \neq k}}^{l_k} \frac{a_i y_i}{s_k} + y_k \\ \left(1 - \frac{1}{s_k}\right)y_{k+1} + y_k \\ \vdots \\ \left(1 - \frac{1}{s_k}\right)y_{l_k} + y_k \end{bmatrix} = \vec{0}.$$

From that, we deduce that

$$\forall i, j \neq k : \left(1 - \frac{1}{s_k}\right)y_i + y_k = 0 = \left(1 - \frac{1}{s_k}\right)y_j + y_k,$$

and hence,  $y_i = y_j$ . Every eigenvector is determined up to any non-zero multiple, so without loss of generality,  $y_i = 1$ . Note that this conclusion depends on the meaningful assumption that  $l_k > 1$  and hence  $s_k > 1$ , but if it is not satisfied, then there is no generalized eigenvalue  $\lambda = 2 - \frac{1}{s_k}$ . From the  $k$ -th row we deduce that

$$y_k = -\sum_{\substack{i=1 \\ i \neq k}}^{l_k} \frac{a_i y_i}{s_k} = -\sum_{\substack{i=1 \\ i \neq k}}^{l_k} \frac{a_i}{s_k} = \frac{1}{s_k} - 1,$$



because  $\sum_{i=1}^{l_k} a_i = s_k$  and  $a_k = 1$ . This completes the proof.  $\square$

**Lemma 4.** Any two eigenvectors  $\vec{x}, \vec{y}$  of the matrix  $L$ , corresponding to different generalized eigenvalues  $\lambda, \nu$ , are orthogonal w.r.t.  $D$ , i.e.,

$$\vec{x}^\top D \vec{y} = 0.$$

*Proof.* It holds:

$$L\vec{x} = \nu D\vec{x} \quad \text{and} \quad L\vec{y} = \lambda D\vec{y},$$

implying that

$$\lambda \vec{x}^\top D \vec{y} = \vec{x}^\top L \vec{y} = \vec{y}^\top L \vec{x} = \nu \vec{y}^\top D \vec{x} = \nu \vec{x}^\top D \vec{y},$$

because the matrices  $L$  and  $D$  are symmetric and all expressions above are 1-dimensional. Utilizing the assumption  $\lambda \neq \nu$ , we deduce that  $\vec{x}^\top D \vec{y} = 0$ .  $\square$

We demonstrate the statement in Lemma 4 by the following valid equalities where  $\vec{y}_1 = \vec{1}$ ,  $\vec{y} \in S_2$  and  $\vec{y}_3$  is in (10) :

$$\begin{aligned} 1. \quad & \vec{y}^\top D \vec{1} = \sum_{\substack{i=1 \\ i \neq k}}^{l_k} a_i y_i = 0, \\ 2. \quad & \vec{y}^\top D \vec{y}_3 = \sum_{\substack{i=1 \\ i \neq k}}^{l_k} a_i y_i = 0. \end{aligned}$$

If we take an arbitrary basis of the space  $S_2$ , we get the set of  $l_k - 2$  eigenvectors corresponding to the generalized eigenvalue 1. They can be used for the realization of the embedding algorithm.

**Theorem 2.** For the fixed component  $G_k$ , the set of eigenvectors of the matrix  $L$ , orthogonal w.r.t.  $D$ , is

$$\text{formed by the vector } \vec{1}, \text{ the vector } \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \frac{1}{s_k} - 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ and by the}$$

corresponding basis of the eigenspace

$$S_2 = \left\{ \vec{y} \in \mathbb{R}^{l_k} \left| \sum_{i=1}^{l_k} a_i y_i = 0 \quad \& \quad y_k = 0 \right. \right\}.$$

The proof follows from the lemmas 2 and 3.

### 4.3 Basis of eigenvectors

Due to the fixed representation of eigenvectors, corresponding to the least and the greatest generalized eigenvalues of the problem in (2), the selection of a basis reduces to the space  $S_2$  (see (9)) – the eigenspace determined by eigenvalue  $\lambda = 1$ . Below, we propose one meaningful example.

For simplicity, we will use the index  $k$  with respect to the order of the points covered by the given partition unit (that is  $k = \frac{l_k+1}{2}$ ). Let the basic function  $A_k$  be covering a set of points  $P = \{p_1, \dots, p_{2k-1}\} = \{p_1, \dots, p_{l_k}\}$  with the node  $p_k = x_k$  and with the symmetric distribution of the data points, that is

$$\forall i = 1, \dots, k-1: p_{k-i} - p_k = p_k - p_{k+1},$$

let the function  $A_k(x)$  be generated be an even function, that is

$$\forall h \in [0, p_k - p_0]: A_k(x_k - h) = A_k(x_k + h),$$

where  $p_0 = x_{k-1}$  is the preceeding node in the primary partition, let the shape of the basic function and the distribution of the data points  $P$  satisfy

$$\forall i = 1, \dots, 2k-2: a_{i+1} - a_i = \begin{cases} \frac{1}{k} & i \leq k-1 \\ -\frac{1}{k} & i \geq k \end{cases},$$

where  $a_i = A_k(p_i)$  and let the function  $A_k(x)$  be strictly increasing on  $[p_0, p_k]$  and strictly decreasing on  $[p_k, p_{2k}]$ , where  $p_{2k} = x_{k+1}$  is the successive node in the primary partition. Then an eigenvector

$$\vec{y} = \begin{bmatrix} 1 - \frac{1}{k} \\ 1 - \frac{2}{k} \\ \vdots \\ \frac{1}{k} \\ 0 \\ -\frac{1}{k} \\ \vdots \\ -1 + \frac{1}{k} \end{bmatrix} = \begin{bmatrix} \frac{l_k-1}{l_k+1} \\ \frac{l_k+1}{l_k+1} \\ \vdots \\ \frac{2}{l_k+1} \\ 0 \\ -\frac{2}{l_k+1} \\ \vdots \\ \frac{1-l_k}{l_k+1} \end{bmatrix},$$

corresponds to the generalized eigenvalue  $\lambda = 1$  and thus belongs to the space  $S_2$ . This can be easily seen from

$$\forall i = 0, \dots, k-1: a_{k-i} = a_{k+i} \quad \& \quad y_{k-i} = -y_{k+i}.$$

We observe that this vector is a discrete version of a line segment, that is its coordinates form a part of an arithmetic sequence with the difference  $\frac{1}{k} = \frac{2}{l_k+1}$ . Note that the conditions above ensure that  $s_k = \sum_{i=1}^{2k-1} a_i = k = \frac{l_k+1}{2}$ . This particular eigenvector agrees with the second polynomial basis function

$P_k^1$  of  $L_2^m(A_k)$ . This example confirms our conjecture regarding the F-transform realization of a dimensionality reduction technique.

Because the spectrum of the matrix  $L$ , corresponding to  $P$ , consists of the values 0, 1 (with the multiplicity  $l_k - 2$ ) and  $\frac{2k-1}{k} = \frac{2l_k}{l_k+1}$ , the set of all eigenvectors, spanning the space  $\mathbb{R}^{l_k}$ , can be written in the form

$$\vec{1}, \begin{bmatrix} l_k - 1 \\ l_k - 3 \\ \vdots \\ 2 \\ 0 \\ -2 \\ \vdots \\ 1 - l_k \end{bmatrix}, \dots \text{ and } \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \frac{1-l_k}{l_k+1} \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

## 5 Conclusion

By the proposed contribution, we justified the fact that the direct F-transform technique is a part of the data reduction process. We gave an example of the graph Laplacian generalized eigenvector basis that can be used for the dimensionality reduction embedding. This example corresponds to the choice of the orthogonal polynomial basis, used by the  $F^1$ -transform.

Beside the scope of the pure dimensionality reduction, the presented theory can also be used in the area of image processing, namely in the image segmentation problem, as it was proposed e.g. by Shi and Malik.

### Acknowledgement

The work of Irina Perfilieva has been partially supported by the project “LQ1602 IT4Innovations excellence in science” and by the Grant Agency of the Czech Republic (project No. 18-06915S).

## References

- [1] M. Belkin, and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6), 1373-1396 (2003)
- [2] I. Perfilieva, Fuzzy Transform: theory and application. *Fuzzy Sets and Systems* 157, 993-1023 (2006)
- [3] I. Perfilieva, M. Daňková, and B. Bede, Towards a higher degree F-transform. *Fuzzy Sets and Systems* 180, 3-19 (2011)
- [4] I. Perfilieva, Dimensionality Reduction by Fuzzy Transforms with Applications to Mathematical Finance. In: Anh L., Dong L., Kreinovich V.,

Thach N. (eds) *Econometrics for Financial Applications*. ECONVN 2018. *Studies in Computational Intelligence*, vol 760. Springer, Cham

- [5] J. Janeček, and I. Perfilieva, F-transform and Dimensionality Reduction: Common and Different. In: *Advances in Intelligent Systems and Computing*, AGOP 2019, Springer, 2019, to appear