Eigenvalue-Based Block Diagonal Representation and Application to p-Nearest Neighbor Graphs

Aylin Taştan, Michael Muma and Abdelhak M. Zoubir
Signal Processing Group
Technische Universität Darmstadt
64283 Darmstadt, Germany
{atastan,muma,zoubir}@spg.tu-darmstadt.de

***Abstract*—Block diagonal structure of the affinity matrix is advantageous, e.g. in graph-based cluster analysis, where each block corresponds to a cluster. However, constructing block diagonal affinity matrices may be challenging and computationally demanding. We propose a new eigenvalue-based block diagonal representation (EBDR) method. The idea is to estimate a block diagonal affinity matrix by finding an approximation to a vector of target eigenvalues. The target eigenvalues, which follow the ideal block-diagonal model, are efficiently determined based on a vector derived from the graph Laplacian that represents the blocks as a piece-wise linear function. The proposed EBDR shows promising performance compared to four optimally tuned state-of-the-art methods in terms of clustering accuracy and computation time using real-data examples.**

***Index Terms*—Block diagonal representation, affinity matrix, similarity matrix, eigenvalues, subspace clustering**

I. Introduction

The construction of an informative graph model plays a crucial role to learn the intrinsic relationships hidden in data and it has numerous applications such as in clustering/classification [1–3], subspace learning [3, 4] and semi-supervised learning [4–6]. In cluster analysis, the graph model represents each feature vector as a vertex and describes the association relationships using an affinity matrix in which block diagonal structure is a commonly desired feature [7–11].

Partly motivated by the natural occurrence of block diagonally structured affinity matrices in cluster analysis, block diagonal representation has been the subject of intense scientific research. Sparse representation is one of the most common ways of constructing a block diagonal affinity matrix [7–10]. An alternative way of constructing block diagonal affinity matrices are p-nearest neighbor graphs which are popular due to their computational simplicity [11]. However, a major challenge for all these methods is to determine the level of sparsity, i.e. the number of neighbors or the regularization parameter. The choice of the sparsity level has been researched by analyzing the similarity coefficients' distribution [12], via supervised learning algorithms [13, 14], geometric interpretations [15] and connectedness [16].

To the best of our knowledge, an unsupervised block diagonal representation method that uses the eigenvalues of a block diagonal affinity matrix to deduce the sparsity level has not been proposed in the literature. Therefore, we first analyze the eigenvalues of the Laplacian matrix based on an ideal block diagonal model (Theorem 1). Then, a key idea is to define a vector that represents the blocks as a piece-wise linear function (Corollary 1.1). This enables a graph construction algorithm building upon the piece-wise linear function that estimates the parameters of the unknown target eigenvalue vector. The proposed *eigenvalues-based block diagonal representation (EBDR)* method is applied to p-nearest neighbor graph construction.

The remaining paper is organized as follows. Section II contains our main theoretical results and the problem formulation. Section III details the proposed EBDR method. A performance evaluation in comparison to popular block diagonal representation methods is the subject of Section IV. Finally, conclusions are drawn in Section V.

II. Theoretical Results and Problem Formulation

A. Motivation: Hidden Information in Eigenvalues

Suppose that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{m\times n}$ with m denoting the dimension and n the number of feature vectors can be represented as a graph $G = \{V, E, \mathbf{W}\}$ where V denotes the vertices, E represents the edges, and $\mathbf{W} \in \mathbb{R}^{n\times n}$ is a block zero-diagonal symmetric affinity matrix[1] whose similarity coefficients within the blocks are generated using a similarity measure[2], while the matrix elements outside the blocks are zero. Let $\mathbf{L} \in \mathbb{R}^{n\times n}$ denote the nonnegative definite Laplacian matrix of the generalized eigen-problem

$$\mathbf{L}\mathbf{y}_i = \lambda_i \mathbf{D}\mathbf{y}_i \tag{1}$$

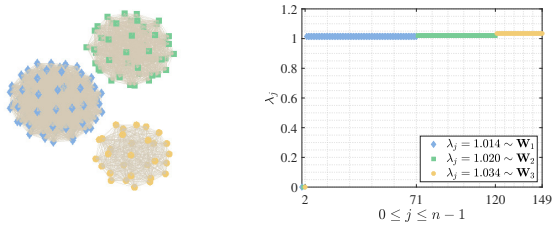
with associated eigenvalues in ascending order given by $0 \leq \lambda_0 \leq \lambda_1 \leq ... \leq \lambda_{n-1}$. Here, $\mathbf{y}_i$ denotes the eigenvector associated with the ith eigenvalue λ_i, $\mathbf{D}\in\mathbb{R}^{n\times n}$ is a diagonal weight matrix with overall edge weights $d_{i,i}=\sum_j w_{i,j}$ on the diagonal and $\mathbf{L}=\mathbf{D}-\mathbf{W}$.

Theorem 1. *Let $\mathbf{W} \in \mathbb{R}^{n\times n}$ be a k block zero-diagonal symmetric affinity matrix with blocks $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_k$ on its diagonal. Each block $\mathbf{W}_i, i=1, ..., k$ is associated to a number*

The authors are with the Signal Processing Group, Technische Universität Darmstadt, Darmstadt, Germany (e-mail: atastan,muma,@spg.tu-darmstadt.de; muma@spg.tu-darmstadt.de; zoubir@spg.tu-darmstadt.de).

[1] A sparse matrix can be transformed into a block diagonal form using the Reverse Cuthill-McKee (RCM) algorithm [17].

[2] E.g. for cosine similarity $w_{i,j} = \mathbf{x}_i^\top \mathbf{x}_j, i \neq j$ s.t. $\|\mathbf{x}_i\|_2 = 1, \|\mathbf{x}_j\|_2 = 1$ where $\|.\|_2$ denotes the ℓ_2 norm.



(a) $G = \{V, E, \mathbf{W}\}$ (b) $\boldsymbol{\lambda}$ associated with Eq. (1)

Fig. 1: Examplary illustration of Theorem 1 ($\mathbf{n}=[70,50,30]^\top$).

$n_i \in \mathbb{Z}^+ > 1$ *of feature vectors. Assuming that each block is concentrated around a similarity constant* $w_i \in \mathbb{R}^+, i=1,...,k$ *with negligibly small variations, the smallest eigenvalue associated with the* i*th block is equal to zero, i.e.* $\lambda_0^{(i)}=0$ *and the remaining* n_i-1 *number of eigenvalues are* $\lambda_{1,...,n_i-1}^{(i)} = \frac{n_i}{n_i-1}$ *where the eigenvalues are in ascending order for the* i*th block such that* $\lambda_0^{(i)} \le \lambda_1 \le ... \le \lambda_{n_i-1}^{(i)}$.

Proof. See Appendix A. □

A synthetic graph model with a $k = 3$ block zero-diagonal affinity matrix and corresponding vector of eigenvalues is shown in Fig. 1. By definition, each block is assumed to be concentrated around a constant $w_i \in \mathbb{R}^+$, e.g. $\mathbf{w}=[0.9, 0.6, 0.3]^\top$. Fig. 1b confirms the results of Theorem 1 that for each block $i=1,...,3$ the smallest eigenvalue is zero and the remaining n_i-1 eigenvalues are $\frac{n_i}{n_i-1}$.

From Theorem 1, it becomes clear that the eigenvalues contain the block size information. This valuable knowledge shall be later used to learn the structure of $\mathbf{W}$ based on the eigenvalues. In the following, a further analysis is performed by defining the vector $\mathbf{v}=[v_1, v_2, ..., v_n]^\top \in \mathbb{R}^n$, such that $v_i = \sum_{j=i}^{n} l_{i,j}, i=1,...,n$ where $l_{i,j}$ is the i,jth element of $\mathbf{L}$.

Corollary 1.1. The vector $\mathbf{v}$ associated with the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$, is a piece-wise linear function, i.e.

$$v_j = f(j) = \begin{cases} (j-\ell_1)w_1 & \text{if} \quad \ell_1 \le j \le u_1 \\ \vdots & \\ (j-\ell_k)w_k & \text{if} \quad \ell_k \le j \le u_k \end{cases}$$

where $\ell_1=1$, $u_1=n_1$, $\ell_i = \sum_{l=1}^{i-1} n_l + 1$ and $u_i = \sum_{l=1}^{i} n_l$ for $i=2,...,k$.

Proof. See Appendix B. □

An illustration of Corollary 1.1 is provided in Fig. 2 for an example consisting of $k = 3$ blocks. The changepoints of $\mathbf{v}$ define the blocks sizes and the coefficients around which the blocks are concentrated. Consequently, $\mathbf{v}$ provides substantial information about the eigenvalues of $\mathbf{L}$, which shall be used to design eigenvalue-based affinity matrix estimation methods that may be computed efficiently through the optimization in a vector space (see Sec. III).

B. Problem Formulation

Given a dataset of feature vectors $\mathbf{X} \in \mathbb{R}^{m \times n}$ and the number of blocks k, the goal of this work is to efficiently estimate a k block zero-diagonal symmetric affinity matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ using the eigenvalue information from Theorem 1 and Corollary 1.1.

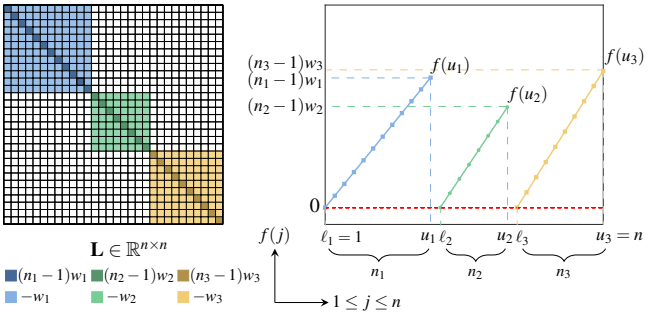


Fig. 2: Examplary plot of $\mathbf{L}$ and $\mathbf{v}$ for $k=3$ blocks.

III. Proposed Method

A. The Optimization Problem

This section proposes a method to represent the data matrix $\mathbf{X}$ as a weighted graph G by finding a k block zero-diagonal affinity matrix $\mathbf{W}$ whose non-zero components in the ith row/column denote the neighbors of the ith node. In principle, if there exists a k block zero-diagonal affinity matrix $\mathbf{W}$ as in Theorem 1, the eigenvalues associated with the Laplacian matrix $\mathbf{L}$ will be in the following form

$$\boldsymbol{\lambda}_o = \text{sort}\Big(\underbrace{0,...,0}_{k}, \underbrace{\frac{n_1}{n_1-1},...,\frac{n_1}{n_1-1}}_{n_1-1}, ..., \underbrace{\frac{n_k}{n_k-1},...,\frac{n_k}{n_k-1}}_{n_k-1}\Big), \quad (2)$$

where $\text{sort}(.)$ denotes sorting operation in ascending order. According to Eq. (2), the estimation of $\mathbf{W}$ can be cast as the following eigenvalue-based optimization program

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{(\hat{p})} = \underset{p_i \in \mathbf{p}}{\text{argmin}} \|\boldsymbol{\lambda}^{(p_i)} - \boldsymbol{\lambda}_o^{(p_i)}\|_2^2. \quad (3)$$

Here, $\hat{\boldsymbol{\lambda}}$ is the estimated vector of eigenvalues which is a function of the estimated number of neighbors (i.e., $\hat{\boldsymbol{\lambda}}=\boldsymbol{\lambda}^{(\hat{p})}$). The estimate is the minimizer of Eq. (1), where p_i is the ith candidate of neighbors from a given vector of candidates $\mathbf{p}=[p_1, p_2, ..., p_{N_p}] \in \mathbb{Z}^{N_p}$. The associated affinity, overall edge weight and Laplacian matrices of dimension $\mathbb{R}^{n \times n}$ are denoted, respectively, by $\mathbf{W}^{(p_i)}$, $\mathbf{D}^{(p_i)}$ and $\mathbf{L}^{(p_i)}$. Finally, $\boldsymbol{\lambda}_o^{(p_i)} \in \mathbb{R}^n$ is the target vector of eigenvalues associated with $\mathbf{L}^{(p_i)}$ whose estimation is detailed in the following section.

B. Estimation of the Target Eigenvalue Vector $\boldsymbol{\lambda}_o^{(p_i)}$

1) Initialization : Possible Block Sizes

Suppose that $\mathbf{v}^{(p_i)} \in \mathbb{R}^n$ denotes the vector $\mathbf{v}$ associated with $\mathbf{W}^{(p_i)}$. Further, let $N_c^{(p_i)} \in \mathbb{Z}^+$ denote the number of changepoints and let $\boldsymbol{\tau}^{(p_i)} = [\tau_1^{(p_i)}, \tau_2^{(p_i)}, ..., \tau_{N_c}^{(p_i)}]^\top \in \mathbb{Z}^+$ be the vector containing corresponding locations in $\mathbf{v}^{(p_i)}$ where $\tau_0^{(p_i)}=0$ and $\tau_{N_c+1}^{(p_i)}=n$. Then, the changepoints in $\mathbf{v}^{(p_i)}$ are detected by minimizing the following penalized least-squares function [18]

$$\sum_{r=1}^{N_c^{(p_i)}+1} \sum_{s=\tau_{r-1}^{(p_i)}+1}^{\tau_r^{(p_i)}} (v_s^{(p_i)} - \hat{v}_s^{(p_i)})^2 + \beta N_c^{(p_i)}, \quad (4)$$

where β is a penalty parameter, $v_s^{(p_i)}$ and $\hat{v}_s^{(p_i)}$ are the sth point in the rth linear piece of $\mathbf{v}^{(p_i)}$ and the corresponding least-squares linear fit $\hat{\mathbf{v}}^{(p_i)}$, respectively. If the decrease in

residual error is smaller than β, Eq. (4) rejects including additional changepoints while all possible changepoints are considered for $\beta = 0$. For a defined maximum number of changepoints $N_{c_{\max}} \in \mathbb{Z}^+$, which is a reasonably small number satisfying $k-1 \leq N_{c_{\max}}$, β is increased gradually as long as the function finds fewer number of changepoints than $N_{c_{\max}}$. Accordingly, a matrix $\mathbf{N}_{\text{cand}}^{(p_i)} = [\mathbf{n}_1^{(p_i)}, \mathbf{n}_2^{(p_i)}, ..., \mathbf{n}_\xi^{(p_i)}]^\top \in \mathbb{R}^{\xi \times k}$ whose rows denote the candidate size vectors is designed by combination of all possible size vectors with $\xi = \binom{N_c}{k-1}$. In practice, the candidate size vectors consisting the block sizes that are smaller than a defined minimum number of nodes in the blocks $n_{\min}$ can be removed from $\mathbf{N}_{\text{cand}}^{(p_i)}$.

2) Plane-based Piece-wise Linear Fit $\mathbf{v}^{(p_i)}$

Suppose that n_l denotes the size of the lth piece from a candidate vector of sizes $\mathbf{n}_{\text{cand}}^{(p_i)} = [n_1^{(p_i)}, n_2^{(p_i)}, ..., n_k^{(p_i)}]^\top \in \mathbb{R}^k$, $\text{cand} = 1, ..., \xi$. Let $\mathbf{S}_l^{(p_i)} = [\mathbf{s}_{l_1}^{(p_i)}, \mathbf{s}_{l_2}^{(p_i)}, ..., \mathbf{s}_{l_{n_l}}^{(p_i)}]^\top \in \mathbb{R}^{n_l \times 2}$ denote a sample matrix associated with the lth linear piece such that $\mathbf{s}_{l_j}^{(p_i)} = [j, v_{l_j}^{(p_i)}]^\top \in \mathbb{R}^2, j = 1, ..., n_l$. Then, the goal of this step is to approximate $\mathbf{v}^{(p_i)}$ using a piece-wise linear function that is determined by estimating k planes, i.e.

$$\hat{\mathcal{P}}_l^{(p_i)} = \{\mathbf{s}_{l_j}^{(p_i)} | \mathbf{s}_{l_j}^{(p_i)} \in \mathbb{R}^2, (\hat{\boldsymbol{\vartheta}}_l^{(p_i)})^\top \mathbf{s}_{l_j}^{(p_i)} + \hat{b}_l^{(p_i)} = 0\}, l = 1, ..., k, \quad (5)$$

where $\hat{\boldsymbol{\vartheta}}_l^{(p_i)} \in \mathbb{R}^2$ and $\hat{b}_l^{(p_i)} \in \mathbb{R}$ denote, respectively, the normal vector and the bias associated with the estimated lth plane $\hat{\mathcal{P}}_l^{(p_i)}$. The estimation can be performed by solving the k individual ordinary eigenvalue problems as in [19]

$$\boldsymbol{\Sigma}_l^{(p_i)} \hat{\boldsymbol{\vartheta}}_l^{(p_i)} = \Lambda_l^{(p_i)} \hat{\boldsymbol{\vartheta}}_l^{(p_i)} \qquad l = 1, ..., k \quad (6)$$

and

$$\hat{b}_l^{(p_i)} = -(\hat{\boldsymbol{\vartheta}}_l^{(p_i)})^\top \boldsymbol{\mu}_l^{(p_i)} \qquad l = 1, ..., k, \quad (7)$$

where $\Lambda_l^{(p_i)} \in \mathbb{R}$ is the smallest eigenvalue associated with the lth plane, and $\boldsymbol{\Sigma}_l^{(p_i)} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{\mu}_l^{(p_i)} \in \mathbb{R}^2$ are, respectively, the covariance matrix and the mean vector of $\mathbf{S}_l^{(p_i)}$.[3] Then, using the estimated parameters of the k planes, each piece in the vector $\mathbf{v}^{(p_i)}$ is estimated as follows

$$(\hat{\boldsymbol{\vartheta}}_l^{(p_i)})^\top \begin{bmatrix} j \\ \hat{v}_{l_j}^{(p_i)} \end{bmatrix} + \hat{b}_l^{(p_i)} = 0, \; l = 1, ..., k, \; j = 1, ..., n_l^{(p_i)}, \quad (8)$$

where $\hat{v}_{l_j}^{(p_i)}$ denotes the jth estimated point in the lth piece. Assuming that for each $\mathbf{n}_{\text{cand}}^{(p_i)} \in \mathbf{N}_{\text{cand}}^{(p_i)}, \text{cand} = 1, ..., \xi$ there exists a piece-wise linear function, the size vector is optimized as follows:

$$\hat{\mathbf{n}} = \underset{\mathbf{n}_{\text{cand}}^{(p_i)} = \mathbf{n}_1^{(p_i)}, ..., \mathbf{n}_\xi^{(p_i)}}{\operatorname{argmin}} \|\mathbf{v}^{(p_i)} - \hat{\mathbf{v}}_{\text{cand}}^{(p_i)}\|_2^2, \quad (9)$$

where $\hat{\mathbf{n}}$ denotes the estimated block size vector and $\hat{\mathbf{v}}_{\text{cand}}^{(p_i)} \in \mathbb{R}^n$ is the estimate of vector $\mathbf{v}^{(p_i)}$ associated with $\mathbf{n}_{\text{cand}}^{(p_i)}$. The proposed EBDR method is summarized for the p-nearest neighbor graphs in Algorithm 1.

[3]The optimal solution to the plane-based piece-wise linear fit problem can be uniquely determined by k covariance matrices and means of the corresponding k clusters (k blocks for our case) when the objective function reaches the optimum. For a detailed information, see Corollary 1-2 in [19].

Algorithm 1: p-nearest Neighbor Graph Construction

Input: $\mathbf{X} \in \mathbb{R}^{m\times n}$, $\mathbf{p} \in \mathbb{R}^{N_p}$, $N_{c_{\max}}$, $n_{\min}$(optional)
for $p_i = p_1, p_2, \ldots, p_{N_p}$ **do**
  Eigenvalue vector $\boldsymbol{\lambda}^{(p_i)}$
  Compute $\mathbf{W}^{(p_i)} \in \mathbb{R}^{n\times n}$ s.t. $w_{i,j} = \mathbf{x}_i^\top \mathbf{x}_j$
  Compute $\boldsymbol{\lambda}^{(p_i)} \in \mathbb{R}^n$ using Eq. (1)
  Target Eigenvalue Vector Estimation $\boldsymbol{\lambda}_{\text{o}}^{(p_i)}$
  Initialization: Possible block sizes
  Compute $\mathbf{N}_{\text{cand}}^{(p_i)} \in \mathbb{R}^{\xi\times k}$ using Eq. (4)
  Plane-based Piece-wise Linear Fit
  for $\mathbf{n}_{\text{cand}}^{(p_i)} = \mathbf{n}_1^{(p_i)}, ..., \mathbf{n}_\xi^{(p_i)}$ **do**
    for $l = 1, ..., k$ **do**
      Calculate $\boldsymbol{\Sigma}^{(p_i)} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{\mu}^{(p_i)} \in \mathbb{R}^2$ for $\mathbf{S}_l^{(p_i)}$
      Find $\hat{\boldsymbol{\vartheta}}_l^{(p_i)} \in \mathbb{R}^2$ and $\hat{b}_l^{(p_i)} \in \mathbb{R}$ via Eq. (6)-(7)
      Substitute $\hat{\boldsymbol{\vartheta}}_l^{(p_i)}, \hat{b}_l^{(p_i)}$ in Eq. (8) to find $\hat{\mathbf{v}}_l^{(p_i)} \in \mathbb{R}^{n_l^{(p_i)}}$
    end
    Form $\hat{\mathbf{v}}_{\text{cand}}^{(p_i)} = [(\hat{\mathbf{v}}_1^{(p_i)})^\top, (\hat{\mathbf{v}}_2^{(p_i)})^\top, ..., (\hat{\mathbf{v}}_k^{(p_i)})^\top]^\top \in \mathbb{R}^n$
  end
  Estimate the block size vector $\mathbf{n}$ using Eq. (9)
  Design $\boldsymbol{\lambda}_{\text{o}}^{(p_i)}$ using Eq. (2)
end
Compute $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{\hat{p}}$ using Eq. (3) and obtain $\mathbf{W}^{(\hat{p})}$
Output: $G^{(\hat{p})} = \{V, E^{(\hat{p})}, \mathbf{W}^{(\hat{p})}\}$

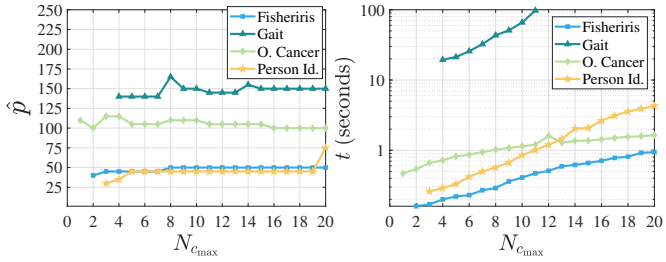
Dataset	k	n	$\mathbf{n}$	$p^\star$	$\hat{p}$
Fisheriris [21]	3	150	$[50, 50, 50]^\top$	50	50
Gait [22, 23]	5	800	$[160, 160, 160, 160, 160]^\top$	160	165
O. Cancer [24]	2	216	$[95, 121]^\top$	100	110
Person Id. [25]	4	187	$[38, 40, 47, 62]^\top$	45	45

TABLE I:Numerical results for real-world datasets ($N_{c_{\max}}$=8).

IV. Experimental Results

In this section, EBDR is benchmarked against three state-of-the art block diagonal representation approaches, i.e. BDSSC [8], BDR-B [9] and IBDLR [10], and a low rank representation method RKLRR [20] that can be reduced to the block diagonal for independent subspaces and the initial matrix containing all neighbors $\mathbf{W}^{n-1}$. The performance of different methods is analyzed in terms of their average clustering accuracy $\bar{p}_{\text{acc}}$ and computation time t using the following real-world datasets: Fisher's iris (Fisheriris) [21], radar-based human gait (Gait) [22, 23], ovarian cancer (O. Cancer) [24] and person identification (Person Id.) [25]. The parameters of the competitors are manually tuned to the best possible $\bar{p}_{\text{acc}}$ by using 500 samples in total. Then, t is summarized for 100 Monte Carlo experiments using the selected parameters. In all experiments, the initial affinity matrix $\mathbf{W}^{n-1}$ is computed using cosine similarity and spectral clustering is performed as partitioning method. EBDR is computed using the following parameters: $n_{\min} = \frac{n}{2k}$, $\mathbf{p} = [5, 10, ..., n-1]$, $N_{c_{\max}} \in [k\text{-}1, ..., 20]$.

In Tab. I, the EBDR application to p-nearest neighbor graphs is benchmarked using real-world datasets. The number of neighbors that provided the best $\bar{p}_{\text{acc}}$ is denoted by $p^\star$. As can be seen, $\hat{p}$ provided similar results to $p^\star$ in all cases. To analyze the effect of $N_{c_{\max}}$ on $\bar{p}_{\text{acc}}$ and t, the estimated nearest neighbor values and computation time are shown for different $N_{c_{\max}}$ values in Fig. 3a and Fig. 3b, respectively. The results



(a) $\hat{p}$ for increasing $N_{c_{\max}}$ (b) t for increasing $N_{c_{\max}}$

Fig. 3: Numerical results for the parameter $N_{c_{\max}}$.

Dataset	$\mathbf{W}^{n-1}$	BDSSC	BDR-B	RKLRR	IBDLR	EBDR
Fisheriris [21]	78.00	96.00	97.33	94.67	94.67	98.00
Gait [22, 23]	79.63	83.13	86.25	86.75	82.75	80.75
O. Cancer [24]	75.00	81.48	86.57	89.35	77.31	79.17
Person Id. [25]	x	95.72	97.33	95.72	95.72	97.33

TABLE II: $\bar{p}_{\text{acc}}(\%)$ for real-world datasets. 'x' denotes the results that produce complex-valued eigenvectors, $N_{c_{\max}}=8$. The numbers indicate the best $\bar{p}_{\text{acc}}$ for the competitors.

Dataset	BDSSC	BDR-B	RKLRR	IBDLR	EBDR($\hat{p}$)	EBDR
Fisheriris [21]	0.174	0.041	0.208	0.573	0.015	0.295
Gait [22, 23]	6.132	4.748	5.394	826.2	0.495	43.29
O. Cancer [24]	2.590	0.099	3.471	1.397	0.041	1.018
Person Id. [25]	0.235	0.489	0.013	0.564	0.019	0.550

TABLE III: t(seconds) for real-world datasets. Except for EBDR the level of sparsity assumed to be known and it is defined as $\hat{p}$ for EBDR($\hat{p}$). $N_{c_{\max}}$=8 in all cases.

show that EBDR approximates $p^\star$ values even for a small number of samples. However, large values of $N_{c_{\max}}$ result in high computational cost, especially in outlier contaminated datasets, e.g. Gait. Lastly, comparisons are drawn in terms of $\bar{p}_{\text{acc}}$ and t for the different methods in Tab. II and Tab. III, respectively.

The clustering accuracy $\bar{p}_{\text{acc}}$ that has been detailed in Tab. II shows the best possible performances for the competitors when the level of sparsity (i.e. the penalty parameter) has been optimally selected. In particular, the competitor clustering accuracy results are the best results according to an oracle selected penalty parameter/s from a grid, while estimating the level of sparsity is part of the optimization for the proposed method. Therefore, Tab. II shows that EBDR improves the performance of $\mathbf{W}^{n-1}$ and performs similar to the best results of the competitors including an unsupervised sparsity parameter estimation p. In terms of t, the proposed method shows a significantly better performance when the level of sparsity is assumed to be known for the competitors. Even when including the nearest neighbor number estimation, EBDR is competitive in terms of speed, which can be further reduced by tuning $\mathbf{p}$, $N_{c_{\max}}$ and $n_{k_{\min}}$.

V. Conclusion and Future Work

The eigenvalues associated with the block affinity matrix are analyzed for the generalized eigen-decomposition to demonstrate the importance of eigenvalues in block affinity matrix design. Based on our theoretical findings on the eigenvalues and the vector $\mathbf{v}$, we proposed EBDR which estimates the number of neighbors by approximating the target eigenvalues. EBDR was benchmarked on different real-world datasets and it showed promising performance compared to four optimally tuned popular approaches in terms of both computation time and the accuracy.

Additional experiments that had to be left out due to space limitations also indicated that the findings in Theorem 1 and Corollary 1.1 may be used to design other fast and unsupervised block diagonal representation methods (e.g., elastic net). Future work will therefore adapt the proposed theory to develop new methods.

Appendix A: Proof of Theorem 1

Let $\mathbf{W}\in\mathbb{R}^{n\times n}$ be a zero-diagonal k block affinity matrix with corresponding Laplacian $\mathbf{L}\in\mathbb{R}^{n\times n}$, i.e.

$$\mathbf{W}=\begin{bmatrix} 0 & w_1 & \dots & w_1 & \dots \\ w_1 & 0 & \dots & w_1 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ w_1 & w_1 & \dots & 0 & \\ & & & & \ddots \\ & & & & & 0 & w_k & \dots & w_k \\ & & & & \dots & w_k & 0 & \dots & w_k \\ & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \dots & w_k & w_k & \dots & 0 \end{bmatrix} \mathbf{L}=\begin{bmatrix} d_1 & -w_1 & \dots & -w_1 & \dots \\ -w_1 & d_1 & \dots & -w_1 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ -w_1 & -w_1 & \dots & d_1 & \\ & & & & \ddots \\ & & & & & d_k & -w_k & \dots & -w_k \\ & & & & \dots & -w_k & d_k & \dots & -w_k \\ & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \dots & -w_k & -w_k & \dots & d_k \end{bmatrix}$$

where $d_i=(n_i-1)w_i, i=1,\dots,k$ and $\mathbf{L}=\mathbf{D}-\mathbf{W}$. To compute the eigenvalues in Eq. (1), $\det(\mathbf{L}-\lambda\mathbf{D})=0$ is considered which can equivalently be written using the determinant properties of block matrices [26], as follows

$$\det(\mathbf{L}-\lambda\mathbf{D})=\det(\mathbf{L}_1-\lambda^{(1)}\mathbf{D}_1)\dots\det(\mathbf{L}_k-\lambda^{(k)}\mathbf{D}_k)=0,$$

where $\mathbf{L}_i\in\mathbb{R}^{n_i\times n_i}$, $\mathbf{D}_i\in\mathbb{R}^{n_i\times n_i}$ and $\lambda^{(i)}$, $i=1,\dots,k$, denote $\mathbf{L}$, $\mathbf{D}$ and λ associated with the ith block, respectively. Further, $\mathbf{L}_i-\lambda^{(i)}\mathbf{D}_i$, $i=1,\dots,k$ can alternatively be written as

$$\underbrace{\begin{bmatrix} c_i & -w_i & \dots & -w_i \\ -w_i & c_i & \dots & -w_i \\ \vdots & \vdots & \ddots & \\ -w_i & -w_i & \dots & c_i \end{bmatrix}}_{\mathbf{L}_i-\lambda^{(i)}\mathbf{D}_i}=\underbrace{\begin{bmatrix} c_i+w_i & 0 & \dots & 0 \\ 0 & c_i+w_i & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & c_i+w_i \end{bmatrix}}_{\mathbf{H}}+\underbrace{\begin{bmatrix} \sqrt{w_i} \\ \sqrt{w_i} \\ \vdots \\ \sqrt{w_i} \end{bmatrix}}_{\mathbf{u}}\underbrace{\begin{bmatrix} -\sqrt{w_i} & -\sqrt{w_i} & \cdots & -\sqrt{w_i} \end{bmatrix}}_{\mathbf{v}^\top}$$

with $c_i=(n_i-1)w_i-\lambda^{(i)}(n_i-1)w_i$. For an invertible matrix $\mathbf{H}\in\mathbb{R}^{n_i\times n_i}$ such that $\mathbf{H}^\dagger=(c_i+w_i)^{-1}\mathbf{I}$, the matrix determinant lemma [27] computes the determinant as $\det(\mathbf{H}+\mathbf{u}\mathbf{v}^\top)=(1+\mathbf{v}^\top\mathbf{H}^\dagger\mathbf{u})\det(\mathbf{H})$ where $\mathbf{u}\in\mathbb{R}^{n_i}$ and $\mathbf{v}\in\mathbb{R}^{n_i}$ are two column vectors. Thus, it holds that

$$\det(\mathbf{L}_i-\lambda^{(i)}\mathbf{D}_i)=\left(1+(-\sqrt{w_i}\mathbf{1})^\top\left(\frac{\sqrt{w_i}}{c_i+w_i}\mathbf{1}\right)\right)(c_i+w_i)^{n_i}$$

where $\mathbf{1}\in\mathbb{Z}^{n_i}$ denotes a column vector of ones. Substituting $\mathbf{1}^\top\mathbf{1}=n_i$ in $\det(\mathbf{L}_i-\lambda^{(i)}\mathbf{D}_i)=0$ leads to

$$\begin{aligned}\left(1-\frac{n_iw_i}{n_iw_i-\lambda^{(i)}(n_i-1)w_i}\right)(n_iw_i-\lambda^{(i)}(n_i-1)w_i)^{n_i}&=0\\ \left(\frac{-\lambda^{(i)}(n_i-1)w_i}{n_iw_i-\lambda^{(i)}(n_i-1)w_i}\right)(n_iw_i-\lambda^{(i)}(n_i-1)w_i)^{n_i}&=0\\ \left(-\lambda^{(i)}(n_i-1)w_i\right)(n_iw_i-\lambda^{(i)}(n_i-1)w_i)^{n_i-1}&=0\end{aligned}$$

For $w_i>0$ and $n_i>1$, the eigenvalues are given by

$$\lambda_0^{(i)}=0 \quad \text{and} \quad \lambda_{1,\dots,n_i-1}^{(i)}=\frac{n_i}{n_i-1}, \qquad i=1,\dots,k.$$

APPENDIX B: PROOF OF COROLLARY 1

By definition, the vector $\mathbf{v}$ is computed by summing up the rows of the upper triangular part of $\mathbf{L}$, i.e.

$$\mathbf{v} = [0, w_1, \ldots, d_1, 0, w_2, \ldots, d_2, \ldots, 0, w_k, \ldots, d_k],$$

where $d_i = (n_i - 1)w_i, i = 1, ..., k$. Since each block includes $n_i \in \{n_1, n_2, ..., n_k\}$ number of nodes the vectors containing lower and upper limits can be defined as follows

$$\boldsymbol{\ell} = \left[1, n_1 + 1, ..., \sum_{i=1}^{k-1} n_i + 1\right] \quad \mathbf{u} = \left[n_1, n_1 + n_2, ..., \sum_{i=1}^{k} n_i\right].$$

Substituting each j in the function $f(j)$ yields the vector:

$$\mathbf{v} = [0, w_1, \ldots, (n_1 - 1)w_1, 0, \ldots, \ldots, 0, w_k, \ldots, (n_k - 1)w_k],$$

which concludes the proof that two vectors are identical.

ACKNOWLEDGMENT

The work of A. Taştan is supported by the Republic of Turkey Ministry of National Education. The work of M. Muma has been funded by the LOEWE initiative (Hesse, Germany) within the emergenCITY centre and is supported by the 'Athene Young Investigator Programme' of Technische Universität Darmstadt, Hesse, Germany.

REFERENCES

[1] A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, pp. 808-828, 2018.

[2] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, pp. 1532-1543, 2018.

[3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, pp. 2765-2781, 2013.

[4] B. Cheng, J. Yang, S. Yan, Y. Fu and T. S. Huang, "Learning with ℓ^1-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, pp. 858-866, 2009.

[5] Q. Liu, Y. Sun, C. Wang, T. Liu and D. Tao, "Elastic net hypergraph learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 26, pp. 452-463, 2016.

[6] X. Zhu, "Semi-supervised learning literature survey," *Comput. Sci.*, vol. 37, pp. 63-67, 2008.

[7] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, pp. 4094-4107, 2012.

[8] J. Feng, Z. Lin, H. Xu and S. Yan, "Robust subspace segmentation with block-diagonal prior," in *Proc. IEEE Conf. Comp. Vision Pattern Recognit.*, pp. 3818-3825, 2014.

[9] C. Lu, J. Feng, Z. Lin, T. Mei and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, pp. 487-501, 2018.

[10] X. Xie, X. Guo, G. Liu and J. Wang, "Implicit block diagonal low-rank representation," *IEEE Trans. Image Process.*, vol. 27, pp. 477-489, 2017.

[11] M. Lucińska and S. T. Wierzchoń, "Spectral clustering based on k-nearest neighbor graph," in *Proc. Int. Conf. Comput. Inf. Syst. Ind. Manage.*, pp. 254-265, 2012.

[12] A. Taştan, M. Muma and A. M. Zoubir, "Sparsity-aware Robust Community Detection," *Signal Process.*, vol. 187, pp. 108147, 2021.

[13] N. García-Pedrajas, J. A. R. Del Castillo and G. Cerruela-García, "A proposal for local k values for k-nearest neighbor rule," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, pp. 470-475, 2015.

[14] S. S. Mullick, S. Datta and S. Das, "Adaptive learning-based k-nearest neighbor classifiers with resilience to class imbalance," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, pp. 5713-5725, 2018.

[15] S. Arora, S. Rao and U. Varizani, "Expander flows, geometric embeddings and graph partitioning" *J. ACM*, vol. 56, pp. 1-37, 2009.

[16] B. Nasihatkon and R. Hartley, "Graph connectivity in sparse subspace clustering," in *Proc. CVPR 2011*, pp. 2137-2144, 2011.

[17] E. Cuthill, and J. McKee, "Reducing the bandwidth of sparse symmetric matrices," in *Proc. 24th Nat. Conf.*, 1969.

[18] R. Killick, P. Fearnhead and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Am. Stat. Assoc.*, vol. 107, pp. 1590-1598, 2012.

[19] X. Yang, H. Yang, F. Zhang, L. Zhang, X. Fan, Q. Ye and L. Fu, "Piecewise linear regression based on plane clustering," *IEEE Access*, vol. 7, pp. 29845-29855, 2019.

[20] S. Xiao, M. Tan, D. Xu and Z. Y. Dong, "Robust kernel low-rank representation", *IEEE Trans. Neural Networks Learn. Syst.*, vol 27, pp. 2268-2281, 2015.

[21] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.

[22] A. -K. Seifert, M. Amin and A. M. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-Doppler signatures," *IEEE Trans. Biomed. Eng.*, vol. 66, pp. 1-11, 2019.

[23] A. Taştan, M. Muma and A. M. Zoubir, "An unsupervised approach for graph-based robust clustering of human gait signatures," in *Proc. 2020 IEEE Radar Conf.*, pp. 1-6, 2020.

[24] T. P. Conrads, V. A. Fusaro, S. Ross, D. Johann, V. Rajapakse, B. A. Hitt, S. M. Steinberg, E. C. Kohn, D. A. Fishman, G. Whitely, J. C. Barrett, L. A. Liotta, E. F. Petricoin and T. D. Veenstra, "High-resolution serum proteomic features for ovarian cancer detection," *Endocrine-related Cancer*, vol. 11, pp. 163-178, 2004.

[25] F. K. Teklehaymanot, A. -K. Seifert, M. Muma, M. G. Amin and A. M. Zoubir, "Bayesian target enumeration and labeling using radar data of human gait," in *Proc. 26th European Signal Process. Conf. (EUSIPCO)*, pp. 1342-1346, 2018.

[26] J. R. Silvester, "Determinants of block matrices", *Math. Gazz.*, vol 84, pp. 460-467, 2000.

[27] J. Ding and A. Zhou, "Eigenvalues of rank-one updated matrices with some applications", *Appl. Math. Lett.*, vol 20, pp. 1223-1226, 2007.