Agglomerative Hierarchical Kernel Spectral Data Clustering

Raghvendra Mall
KU Leuven, ESAT/STADIUS
B-3001 Leuven, Belgium
Email: raghvendra.mall@esat.kuleuven.be

Rocco Langone
KU Leuven, ESAT/STADIUS
B-3001 Leuven, Belgium
rocco.langone@esat.kuleuven.be

Johan A.K. Suykens
KU Leuven, ESAT/STADIUS
B-3001 Leuven, Belgium
johan.suykens@esat.kuleuven.be

***Abstract*—In this paper we extend the agglomerative hierarchical kernel spectral clustering (AH-KSC [1]) technique from networks to datasets and images. The kernel spectral clustering (KSC) technique builds a clustering model in a primal-dual optimization framework. The dual solution leads to an eigen-decomposition. The clustering model consists of kernel evaluations, projections onto the eigenvectors and a powerful out-of-sample extension property. We first estimate the optimal model parameters using the balanced angular fitting (BAF) [2] criterion. We then exploit the eigen-projections corresponding to these parameters to automatically identify a set of increasing distance thresholds. These distance thresholds provide the clusters at different levels of hierarchy in the dataset which are merged in an agglomerative fashion as shown in [1], [4]. We showcase the effectiveness of the AH-KSC method on several datasets and real world images. We compare the AH-KSC method with several agglomerative hierarchical clustering techniques and overcome the issues of hierarchical KSC technique proposed in [5].**

I. Introduction

Clustering is a widely used technique in the field of data mining, machine learning, vector quantization, data compression and many other tasks [6]–[9]. The core concept in clustering is to locate the inherent groups present within the dataset. These groups are defined as clusters when the data within the groups are more similar to one another in comparison to the similarity of the data between the groups.

Spectral clustering methods [10]–[13] are one class of clustering methods. For spectral clustering methods the clustering information is obtained from eigenvectors of a matrix derived from pairwise similarities of the data. The clusters have localized structure in the eigenspace. These clusters are obtained by performing a simple clustering technique (k-means) on the eigenvectors. One of the issues with spectral clustering methods is that they need to construct a $N \times N$ similarity matrix and perform its eigen-decomposition $O(N^3)$. As the size N increases, it becomes computationally unfeasible and memory extensive.

Recently, a kernel spectral clustering (KSC) algorithm based on weighted kernel PCA formulation was proposed in [14]. The KSC method builds a model on a subset of the data in a primal-dual framework. This subset acts as a representative subset of the dataset and is selected by maximizing the quadratic Rènyi entropy criterion proposed in [15], [16]. The model is expressed in the primal as a set of projections in high dimensional feature spaces. The corresponding dual model has a powerful out-of-sample extension property which allows to infer the cluster affiliation for previously unseen data.

During the training stage, the KSC methodology solves an eigen-decomposition problem. The dual clustering model is expressed as linear combinations of kernel evaluations with the eigenvectors used as the coefficients. We then perform the model selection using the balanced angular fitting (BAF) criterion [2], [3]. The BAF criterion provides the optimal kernel parameter σ for the radial basis kernel function used in case of datasets and σ_{χ^2} kernel function used in case of images. However, in case of networks a linear kernel is sufficient. In this paper, we showcase that the BAF criterion can capture multiple peaks corresponding to different levels of hierarchy for an optimal σ. It also identifies the optimal number of clusters k for which the BAF criterion is maximum. This generally corresponds to the highest level of hierarchy.

In order to estimate these peaks and obtain a hierarchical organization for the given dataset we exploit the structure of the eigen-projections for the validation set obtained from the dual clustering model using the optimal kernel parameter (σ/σ_{χ^2}) and the out-of-sample extension property. The concept of exploiting the structure of the eigen-projections was first introduced in [17] for automatically determining the number of communities k in a network. It was then extended to identify multilevel hierarchical organizations in large scale networks in [1], [4]. However, in the case of networks the normalized linear kernel is used [1], [4] and there is no kernel parameter (σ). We used the same concept to obtain a hierarchical structure in a bottom-up fashion after obtaining the optimal σ from the BAF model selection criterion.

A hierarchical kernel spectral clustering technique was proposed in [5]. There the authors used multiple scales of the kernel parameter σ to obtain a KSC model with different k at each level of hierarchy. Then, for each KSC model they obtain the cluster memberships for the full dataset using the out-of-sample extension property. A special linkage criterion was used to determine which clusters at a given level merged based on cluster memberships of the full dataset. One of the issues with this method is that during the merging, there might be some points of the merging clusters which go to a different clusters as mentioned in [5]. These points are forced to join the merging cluster of the majority. Thus, the hierarchical connections between the layers are more ad-hoc. We overcome this problem in the AH-KSC method as it provides a natural agglomerative bottom-up hierarchical organization.

Figure 1 provides an overview of the steps involved in AH-KSC method and Figure 2 depicts the result of AH-KSC approach on a synthetic dataset and Figure 3 showcases the same for a real world image (Id:119082).

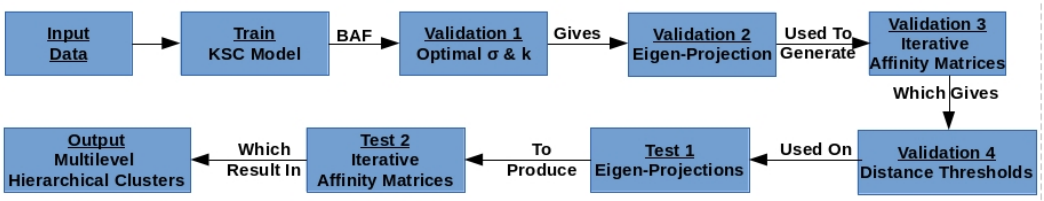
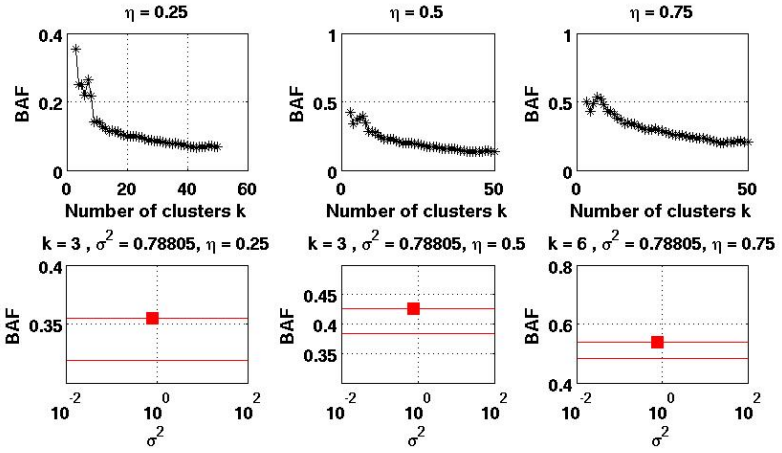
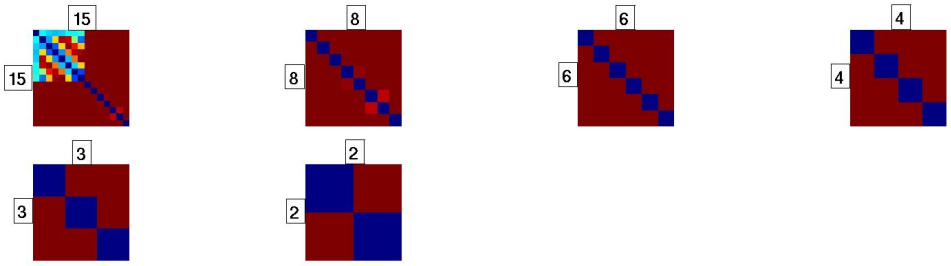


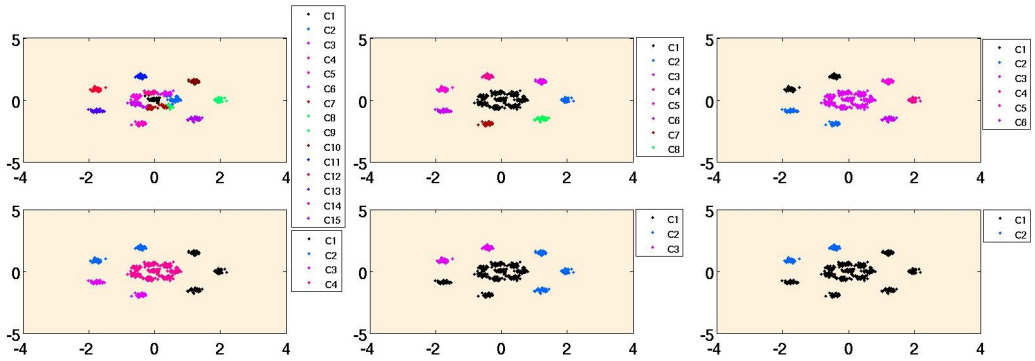
Fig. 1: Steps of AH-KSC method as described in [4] with addition of the step where the optimal σ and k are estimated.



(a) Selection of optimal σ using BAF criterion and illustation of the multiple peaks in the BAF curve for different values of k.



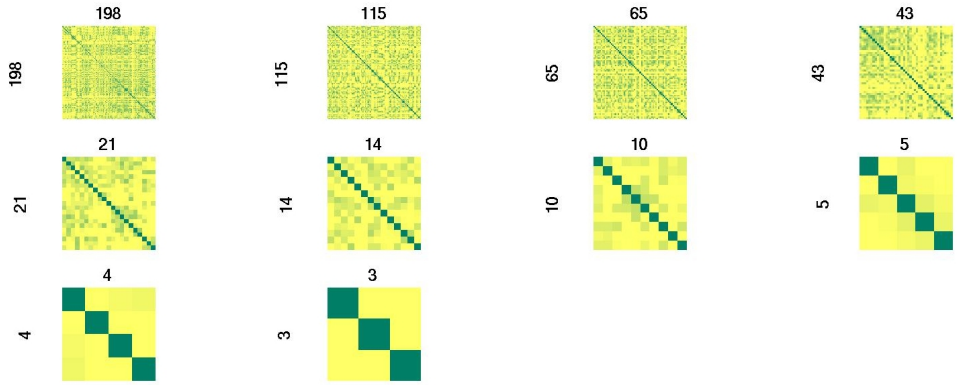
(b) Affinity matrices created at different levels of hierarchy in left to right order. The x, y-axis represent the size of the affinity matrix and the number of clusters k at each level of hierarchy.



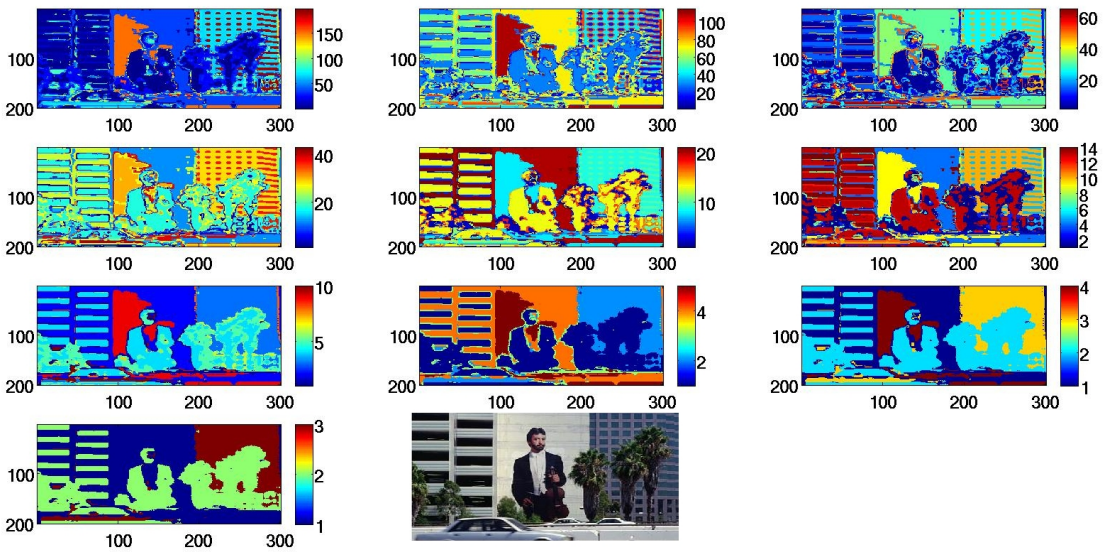
(c) The cluster memberships at different levels of hierarchy

Fig. 2: Result of AH-KSC method on a synthetic dataset comprising 15 Gaussians.

Section II provides a brief description of the KSC method. Section III describes the AH-KSC algorithm. The experimental results are detailed in Section IV. We provide the conclusion in Section V.



(a) Affinity matrices created at different levels of hierarchy from left to right. The x, y-axis represent the size of the affinity matrix and number of clusters k at each level of hierarchy.



(b) The image segments at different levels of hierarchy along with the original image.

Fig. 3: Result of AH-KSC method on a real world image dataset (Id:119082).

II. Kernel Spectral Clustering method

We first provide a brief description of the Kernel Spectral Clustering (KSC) methodology introduced in [14].

A. Primal-Dual Weighted Kernel PCA framework

Given the training set $\mathcal{D} = \{x_i\}_{i=1}^{N_{tr}}$, $x_i \in \mathbb{R}^d$, the training points are selected by maximizing the quadratic Rènyi criterion as depicted in [15], [16] and [18]. Thus, the training model is built on a subset of the entire dataset. Here x_i represents the i^{th} training data point and the training set is represented by X_{tr}. The number of points in the training set is N_{tr}. Given $\mathcal{D}$ and a user-defined maximum number of clusters $maxk$, the primal problem of the spectral clustering via weighted kernel PCA is formulated as follows [14]:

$$\begin{aligned} \min_{w^{(l)},e^{(l)},b_l} \quad & \frac{1}{2}\sum_{l=1}^{k-1} w^{(l)\intercal} w^{(l)} - \frac{1}{2N_{tr}}\sum_{l=1}^{k-1}\gamma_l e^{(l)\intercal} D_\Omega^{-1} e^{(l)} \\ \text{such that} \quad & e^{(l)} = \Phi w^{(l)} + b_l 1_{N_{tr}}, l = 1,\ldots,maxk-1, \end{aligned} \tag{1}$$

where $e^{(l)} = [e_1^{(l)},\ldots,e_{N_{tr}}^{(l)}]^\intercal$ are the projections onto the eigenspace, $l = 1,\ldots,maxk-1$ indicates the number of score variables required to encode the $maxk$ clusters, $D_\Omega^{-1} \in \mathbb{R}^{N_{tr}\times N_{tr}}$ is the inverse of the degree matrix associated to the kernel matrix Ω. Φ is the $N_{tr} \times n_h$ feature matrix, $\Phi = [\phi(x_1)^\intercal;\ldots;\phi(x_{N_{tr}})^\intercal]$ and $\gamma_l \in \mathbb{R}^+$ are the regularization constants. The kernel matrix Ω is obtained by calculating the similarity between each pair of data points in the training set. Each element of Ω, denoted as $\Omega_{ij} = K(x_i,x_j) = \phi(x_i)^\intercal\phi(x_j)$ is obtained for example by using the radial basis function (RBF) kernel. The clustering model in the primal is then represented by:

$$e_i^{(l)} = w^{(l)\intercal}\phi(x_i) + b_l, i = 1,\ldots,N_{tr}, \tag{2}$$

where $\phi : \mathbb{R}^d \to \mathbb{R}^{n_h}$ is the mapping to a high-dimensional feature space n_h, b_l are the bias terms, $l = 1,\ldots,maxk$-1. The dual problem corresponding to this primal formulation is:

$$D_\Omega^{-1} M_D \Omega \alpha^{(l)} = \lambda_l \alpha^{(l)}, \tag{3}$$

where M_D is the centering matrix which is defined as $M_D = \mathrm{I}_{N_{tr}} - (\frac{(1_{N_{tr}} 1_{N_{tr}}^{\intercal} D_{\Omega}^{-1})}{1_{N_{tr}}^{\intercal} D_{\Omega}^{-1} 1_{N_{tr}}})$. The $\alpha^{(l)}$ are the dual variables and the positive definite kernel function $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ plays the role of similarity function. This dual problem is closely related to the random walk model as shown in [14].

B. Out-of-Sample Extension Model

The projections $e^{(l)}$ define the cluster indicators for the training data using the l^{th} eigenvector. In the case of an unseen data point x, the predictive model becomes:

$$\hat{e}^{(l)}(x) = \sum_{i=1}^{N_{tr}} \alpha_i^{(l)} K(x, x_i) + b_l. \tag{4}$$

The validation stage is used to obtain the model parameters like the kernel parameter (σ for RBF kernel or σ_{χ^2} for RBF chi-squared kernel) and the optimal number of clusters k.

C. Model Selection

We use the Balanced Angular Fit (BAF) criterion proposed in [2] for cluster quality evaluation. It is defined as:

$$\begin{aligned} &\mathrm{BAF}(k,\sigma) = \textstyle\sum_{p=1}^{k} \sum_{valid_{(i,\sigma)} \in Q_p} \frac{1}{k} \cdot \frac{MS(valid_{(i,\sigma)})}{|Q_p|} + \eta \frac{\min_l |Q_l|}{\max_m |Q_m|}, \\ &MS(valid_{(i,\sigma)}) = \max_j \cos(\theta_{j,valid_{(i,\sigma)}}), j = 1, \ldots, k \\ &cos(\theta_{j,valid_{(i,\sigma)}}) = \frac{\mu_j^{\intercal} e_{valid_{(i,\sigma)}}}{\|\mu_j\| \|e_{valid_{(i,\sigma)}}\|}, j = 1, \ldots, k. \end{aligned} \tag{5}$$

where $e_{valid_{(i,\sigma)}}$ represents projection of i^{th} validation point for a given σ, μ_j is mean projection of all validation points in cluster j and Q_p represents the set of validation points belonging to cluster p and $|Q_p|$ is its cardinality.

This criterion works on the principle of angular similarity of validation points to the cluster means w.r.t. a given k and σ. Validation points are allocated to the clusters to which (μ_j) they have the least angular distance. We use a regularizer η to vary the priority between angular fitting and balance. We locate the optimal σ over a grid of (σ, k) values by selecting the one for which the BAF value is maximum over this grid. The BAF criterion varies from [-1, 1] and higher values are better for a given value of k.

III. Agglomerative Hierarchical KSC approach

In [4], the authors show that the dual predictive KSC model can be used to obtain the latent variable matrix for the validation set as $P_{valid} = [e_1; \ldots; e_{N_{valid}}]$. The P_{valid} matrix is a $N_{valid} \times (maxk\text{-}1)$ matrix. The test projection set is denoted by P_{test}. In [17], the authors created an affinity matrix S_{valid} using P_{valid} as:

$$S_{valid}(i,j) = CosDist(e_i, e_j) = 1 - \cos(e_i, e_j) = 1 - \frac{e_i^{\intercal} e_j}{\|e_i\| \|e_j\|}, \tag{6}$$

where $CosDist(\cdot,\cdot)$ function calculates the cosine distance between 2 vectors and takes values between $[0, 2]$. Data points which belong to the same cluster have $CosDist(e_i, e_j)$ closer to 0, $\forall i, j$ in the same cluster. It was shown in [17] that a rotation of the S_{valid} matrix has a block diagonal structure where the number of block diagonals is equal to number of clusters (k) in the dataset. In [4] we showed that the affinity matrix generated at one level of hierarchy depends on the affinity matrix of the previous level and a kernel function. This can be depicted in the form of an equation as:

$$\begin{aligned} S^{(h)}(i,j) &= \sum_{m \in C_i^{(h-1)}} \sum_{l \in C_j^{(h-1)}} \kappa(i,j) S^{(h-1)}(m,l) \\ \kappa(i,j) &= \frac{1}{|C_i^{(h-1)}| \times |C_j^{(h-1)}|}, \end{aligned} \tag{7}$$

where $S^{(h)}(i,j)$ represents the $(i,j)^{th}$ element of the affinity matrix at level h of hierarchy. $C_i^{(h-1)}$ represents the set of points in the i^{th} cluster at level h-1 of hierarchy, $|\cdot|$ represents cardinality and $\kappa(i,j)$ is the kernel function which acts as a normalization constant.

$S_{valid}^{(0)}$ is obtained by calculating the $CosDist(\cdot,\cdot)$ between the elements of P_{valid} matrix as shown in equation (6). It was shown in [1] and [17] that $t^{(0)} = 0.15$ is most suitable threshold for the lowest level of hierarchy. A detailed description of the AH-KSC approach can be obtained from [1], [4]. Algorithm 1 summarizes the AH-KSC approach.

Algorithm 1: AH-KSC Method for Data Clustering

Data: Given a dataset $\mathcal{D}$ of data points or images.
Result: Agglomerative Hierarchical KSC of the dataset.

1. Divide data into train,validation and test set.
2. Use $\mathcal{D} = \{x_i\}_{i=1}^{N_{tr}}$, $x_i \in \mathbb{R}^d$ as the training set.
3. Perform KSC on $\mathcal{D}$ to obtain the predictive model as in equation (4).
4. **Use BAF criterion to obtain optimal kernel parameter σ which results in multiple peaks in the BAF curve for different values of k.** `/* Additional Step required for datasets & images. */`
5. Obtain $P_{valid} = [e_1; \ldots; e_{N_{valid}}]$.
6. $S_{valid}^{(0)}(i,j) = CosDist(e_i, e_j) = 1 - \frac{e_i^{\intercal} e_j}{\|e_i\| \|e_j\|}$, $\forall e_i, e_j \in P_{valid}$ using equation (6).
7. Begin validation stage with: $h = 0$, $t^{(0)} = 0.15$.
8. $[C^{(0)}, k] = GreedyMaxOrder(S_{valid}^{(0)}, t^{(0)})$. Add $t^{(0)}$ to the set $\mathcal{T}$ and $C^{(0)}$ to the set $\mathcal{C}$.
9. **while** $k > 1$ **do**
10. $h := h + 1$.
11. Create $S_{valid}^{(h)}$ from $S_{valid}^{(h-1)}$ and $C^{(h-1)}$ using the concept in equation (7).
12. Calculate $t^{(h)} = \mathrm{mean}(\min_j(A_{valid}^{(h)}(i,j))), i \neq j$.
13. $[C^{(h)}, k] = GreedyMaxOrder(S_{valid}^{(h)}, t^{(h)})$.
14. Add $t^{(h)}$ to the set $\mathcal{T}$ and $C^{(h)}$ to the set $\mathcal{C}$.
15. **end**
`/* Iterative procedure to get` $\mathcal{T}$. `*/`
16. Obtain P_{test} like P_{valid} and begin with: $h = 1$, $t^{(1)} \in \mathcal{T}$.
17. $[S_{test}^{(2)}, C^{(1)}, k] = GreedyFirstOrder(P_{test}, t^{(1)})$. Add $C^{(1)}$ to the set $\mathcal{C}$.
18. **foreach** $t^{(h)} \in \mathcal{T}$, $h > 1$ **do**
19. $[C^{(h)}, k] = GreedyMaxOrder(S_{test}^{(h)}, t^{(h)})$.
20. Add $C^{(h)}$ to the set $\mathcal{C}$.
21. Create $S_{test}^{(h+1)}$ from $S_{test}^{(h)}$ & $C^{(h)}$ using equation (7).
22. **end**
23. Obtain the set $\mathcal{C}$ for test set and propagate cluster memberships from 1^{st} to topmost level of hierarchy.

IV. EXPERIMENTS

We conducted experiments on several synthetic datasets available at http://cs.joensuu.fi/sipu/datasets/ and some real world images from http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/. For the synthetic datasets the ground truth or cluster memberships corresponding to flat clustering are known beforehand. Hence, we can use external cluster quality measure like Adjusted Rand Index (ARI) [19] for evaluating the quality of the clusters. However, for these datasets, only flat clustering or cluster membership for one level of hierarchy is known.

We use 30% and 40% of the data as training and validation set. We use the entire dataset as test set. To evaluate the quality of the clusters obtained at multiple levels of hierarchy we utilize internal quality metrics like silhouette (Sil) and Davies-Bouldin index (DB) [19]. Silhouette values are normalized between [0,1] and higher values correspond to better quality clusters while DB has opposite behavior. Lower values of DB close to 0 represent better quality clusters. We define an F-score as ratio between Sil and DB and observe the level of hierarchy for which the F-score is maximum. Table I provides a brief description of the datasets used in the experiments.

Datasets			
Name	Total Points (N)	Dimensions	Optimal k
Aggregation	788	2	7
Dim2	1351	2	9
R15	600	2	15
S1	5000	2	15
Images			
Number	Pixels		Optimal k
Id:10001	6,400		4
Id:10002	6,400		5
Id:100007	154,401		4
Id:119082	154,401		5

TABLE I: Datasets used in the experiments. For synthetic datasets the optimal k is equivalent to ground truth number of clusters. For images optimal k is selected based on the maximum F-score value for SL, CL, ML, WL and AL agglomerative clustering techniques. This optimal k for images is usually not the same as that for AH-KSC method.

A. Experiments on Datasets

We compare AH-KSC method with traditional agglomerative hierarchical clustering techniques which merge clusters based on a linkage measure that specifies how dissimilar 2 clusters are. We compare against linkage techniques [6] including single (SL), complete (CL), median (ML) and weighted linkage (WL). We also compare against average link (AL) technique for datasets. One of the issues with these techniques is that they calculate pairwise similarities and at the lowest level of hierarchy the complexity becomes $O(N^2)$ which is computationally expensive. The AH-KSC method overcomes this problem by obtaining a model on a subset of the data. It then estimates the optimal model parameters (σ) using the BAF criterion following which it performs the agglomerative hierarchical clustering. It uses the $GreedyFirstOrder$ approach [1] which overcomes the issue of pairwise similarity calculation.

For all our experiments we perform 10 randomizations and report the mean results for various quality metrics. In Table II we provide a comparison of the AH-KSC method with other hierarchical clustering techniques w.r.t. the various quality measures like F-score, DB, Sil and ARI. For each method we present the best value for a given quality measure which corresponds to a clustering for that level of hierarchy which has the maximum F-score. From Table II, we observe that the AH-KSC method gives the best results on 3 datasets for F-score and DB measure and 2 datasets for Sil internal quality measure.

Dataset	AH-KSC	SL	AL	CL	ML	WL
	F-score					
Aggregation	**1.34**	0.84	1.08	1.01	0.98	0.92
Dim2	2.27	**7.37**	5.30	4.54	5.40	5.40
R15	**4.75**	0.92	3.32	3.28	3.18	3.15
S1	**2.96**	0.0	1.81	1.55	1.58	1.76
	DB					
Aggregation	**0.466**	0.65	0.64	0.66	0.68	70
Dim2	0.37	**0.12**	0.17	0.20	0.17	0.17
R15	**0.17**	0.81	0.27	0.27	0.28	0.29
S1	**0.29**	1.7	0.45	0.51	0.50	0.46
	Sil					
Aggregation	**0.69**	0.55	**0.69**	0.67	0.66	0.65
Dim2	0.85	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
R15	0.78	0.74	**0.90**	0.90	0.90	0.90
S1	**0.88**	0.54	0.81	0.79	0.79	0.81
	ARI					
Aggregation	**0.94**	0.81	0.93	0.78	0.67	0.71
R15	0.87	0.54	**0.99**	0.98	0.98	0.99

TABLE II: Comparison of AH-KSC method with single, average, complete, median and weight linkage techniques [6] based on various quality measures. The bold represent the best approach among the methods for a given dataset.

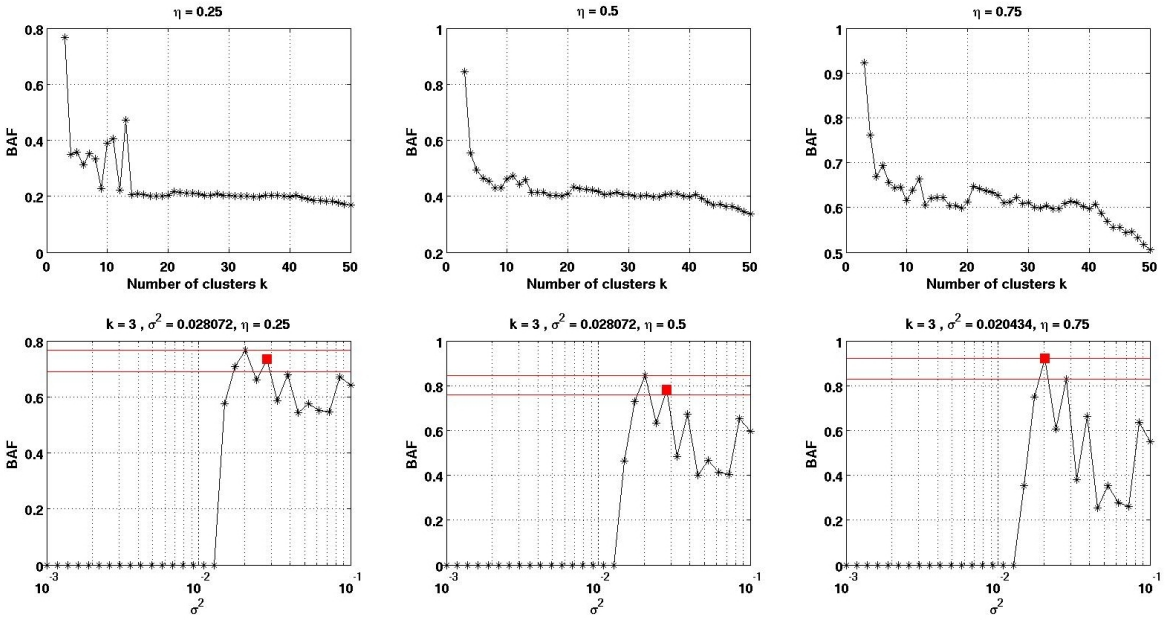
Figure 4 compares AH-KSC method with other clustering techniques for Aggregation dataset. We plot the tree based hierarchical structure for AH-KSC method in Figure 5a. Figure 5 also showcases the dendogram plots for SL, AL, CL, ML and WL hierarchical clustering techniques. The dendogram plots (5b,5c,5d,5e,5f) combine only 2 clusters at a given level of hierarchy. The tree based plot doesn't suffer from this drawback and can combine multiple set of clusters at a given level of hierarchy.

B. Experiments on Images

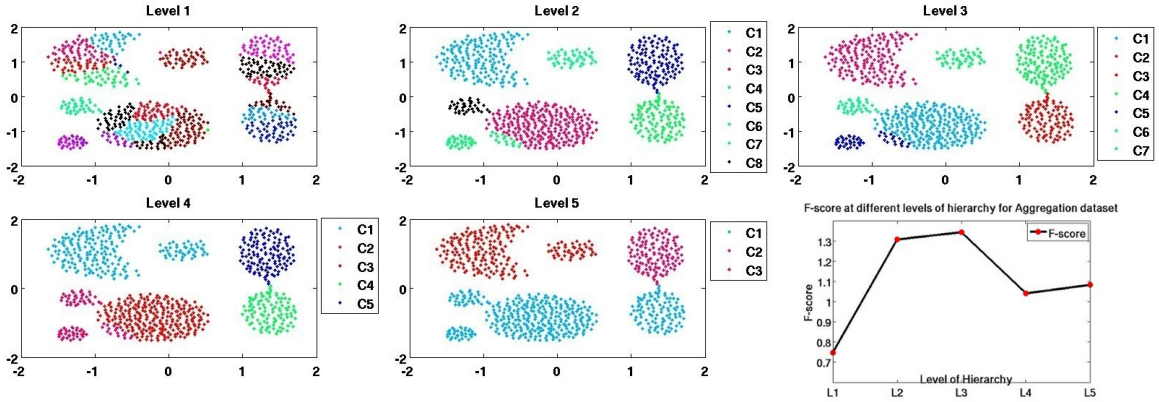
We illustrate segmentation on an image (Id:100007) such that each pixel is transformed into a histogram and the χ^2 distance is used in the RBF kernel with bandwidth σ_{χ^2}.

Figure 6 represents the image segmentation results for AH-KSC method at various level of hierarchy and the results for SL, CL, ML and WL agglomerative clustering techniques for the layer corresponding to which the F-score is highest. For the AH-KSC method we scale the image to 200×250 i.e. $50,000$ pixels. For the other hierarchical clustering techniques the image is scaled down to 100×150 i.e. $15,000$ pixels. This is because otherwise they become memory extensive. The Average Link technique is computationally very expensive and therefore not used during evaluation. From Figure 6a we observe that hierarchical segments obtained by AH-KSC method seem much more natural than those obtained by other clustering techniques as shown in Figure 6b. This is also verified by the highest Sil value for AH-KSC method in comparison to other techniques depicted in Table III.

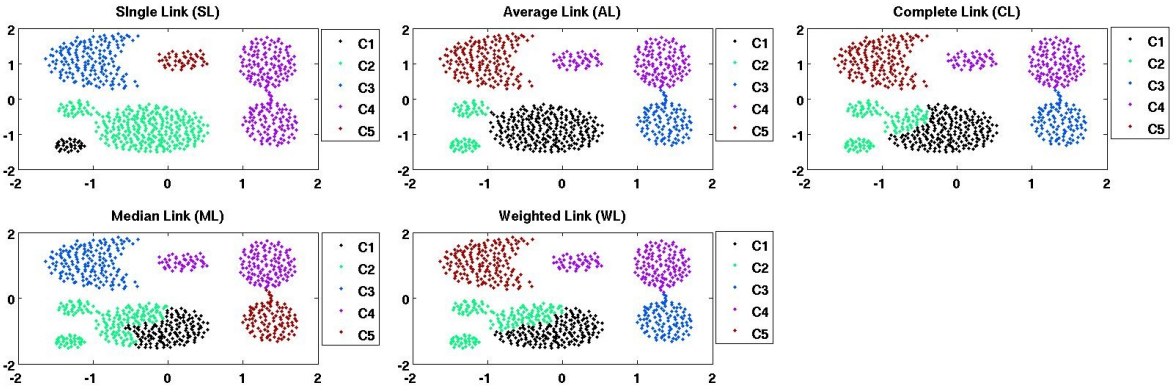
From Table III, we can observe that the AH-KSC method provides the best F-score for 2 images, best DB value for 1 image and best Sil value for 3 images. We observe that for image Id:100007 AH-KSC approach has the best Sil value but the weighted link technique has the lowest DB value which



(a) Optimal σ selection using BAF criterion. η determines the weight of the balance part. The σ which results in highest BAF value for maximum number of times for all η values is selected. In this example, optimal $\sigma = 0.028$ for $k = 3$. $k = 3$ should occur at one level of hierarchy for the AH-KSC method.

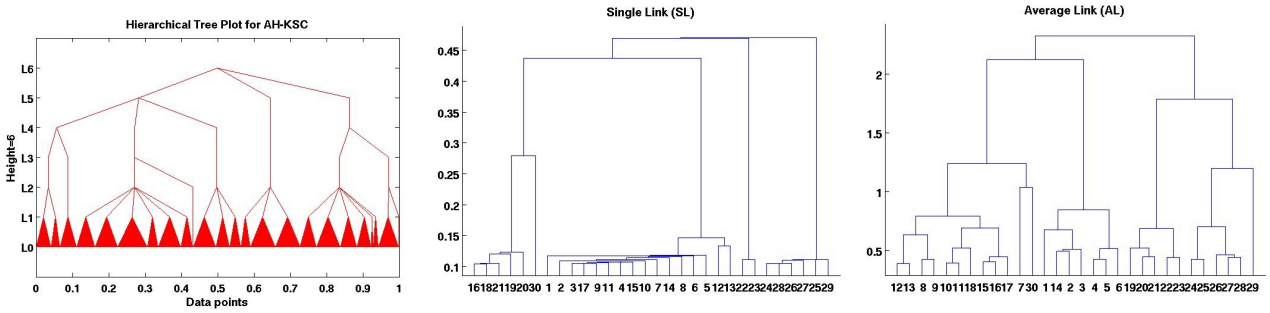


(b) 5 levels of hierarchy produced by AH-KSC method and best F-score at $k = 7$ for Level 3.

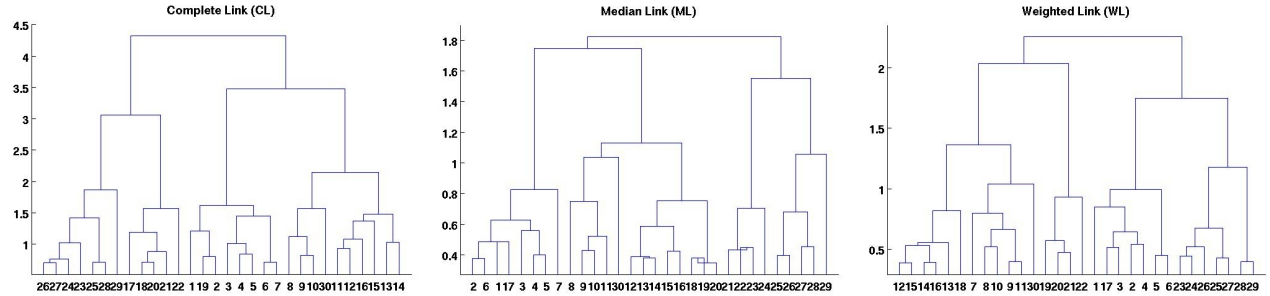


(c) Best results for other hierarchical techniques where F-score is maximum for $k = 5$.

Fig. 4: Results of different hierarchical clustering technique on Aggregation dataset.



(a) Tree plot for AH-KSC method (b) Dendogram for Single Link (c) Dendogram for Average Link



(d) Dendogram for Complete Link (e) Dendogram for Median Link (f) Dendogram for Weighted Link

Fig. 5: Plots of different hierarchical structures for Aggregation dataset. AH-KSC method allows multiple set of clusters to combine at a given level of hierarchcy which is not permitted by the dendogram structure.

is quite small and results in better F-score for WL method in comparison to AH-KSC approach.

Image	AH-KSC	SL	CL	ML	WL
	F-score				
Id:10001	**2.08**	0.394	1.57	1.35	1.35
Id:10002	1.02	0.70	1.37	1.2	**1.40**
Id:100007	0.73	0.38	0.41	0.68	**1.00**
Id:119082	**0.59**	0.57	0.44	0.56	0.59
	DB				
Id:10001	**0.41**	0.68	0.55	0.62	0.62
Id:10002	0.73	0.80	0.57	0.63	**0.55**
Id:100007	0.89	1.02	1.41	0.75	**0.48**
Id:119082	0.97	**0.80**	1.1	0.88	0.81
	Sil				
Id:10001	**0.85**	0.27	0.85	0.84	0.84
Id:10002	0.75	0.56	**0.77**	0.76	0.77
Id:100007	**0.65**	0.39	0.58	0.51	0.48
Id:119082	**0.57**	0.45	0.48	0.49	0.49

TABLE III: Comparison of AH-KSC method with other Agglomerative Hierarchical techniques based on various quality measures for images. The bold represent the best approach among the methods for a given image.

V. Conclusion

We extended the AH-KSC method [1], [4] from networks to data clustering. We obtained the optimal kernel parameter σ during the model selection stage using the BAF criterion. We used this optimal σ to generate the KSC clustering model. We then generated the eigen-projections using the out-of-sample extension property. These were further utilized to estimate a set of distance thresholds ($\mathcal{T}$). We iteratively built affinity matrices from these eigen-projections and used these distance thresholds to obtain an agglomerative hierarchical organization in a bottom-up fashion. We compared the AH-KSC method with other agglomerative hierarchical clustering techniques and showed its effectiveness on several datasets and images.

Acknowledgments

This work was supported by EU: ERC AdG A-DATADRIVE-B (290923), Research Council KUL: GOA/10/-/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants-Flemish Government; FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grants; IWT: projects: SBO POM (100031); PhD/Postdoc grants; iMinds Medical Information Technologies SBO 2014-Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017).

References

[1] R. Mall, R. Langone and J.A.K. Suykens. *Agglomerative Hierarchical Kernel Spectral Clustering for Large Scale Networks*, In Proceedings of European Symposium on Artificial Neural Networks (ESANN), Brugges, Belgium, April, 2014.

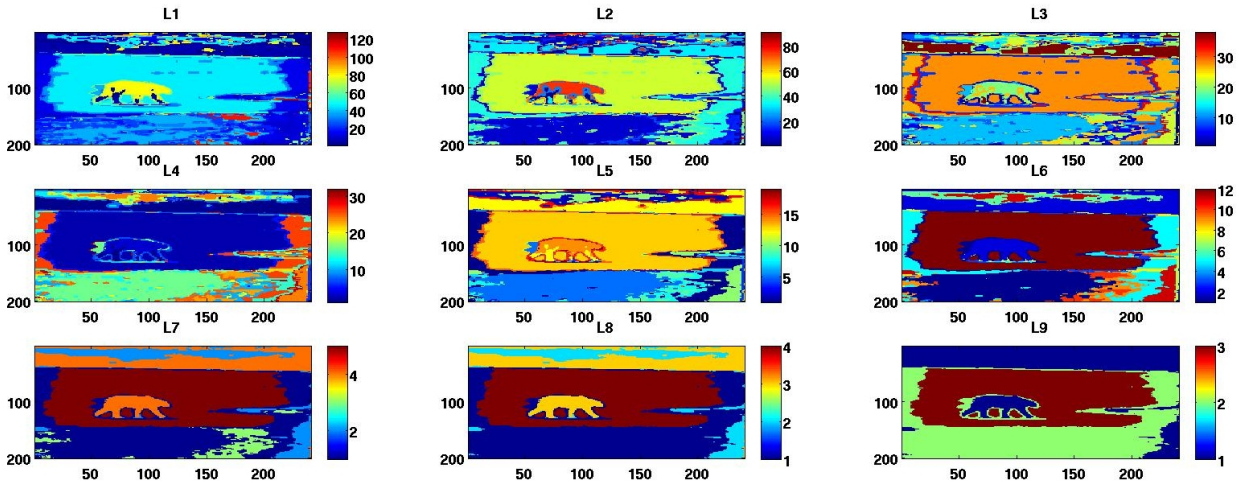
[2] R. Mall, R. Langone and J.A.K. Suykens. *Kernel Spectral Clustering for Big Data Networks*, Entropy (SI: Big Data), 15(5):1567-1586, 2013.

[3] R. Mall, R. Langone and J.A.K. Suykens. *FURS: Fast and Unique Representative Subset selection retaining large-scale community structure.* Social Network Analysis and Mining 3(4):1075-1095, 2013.

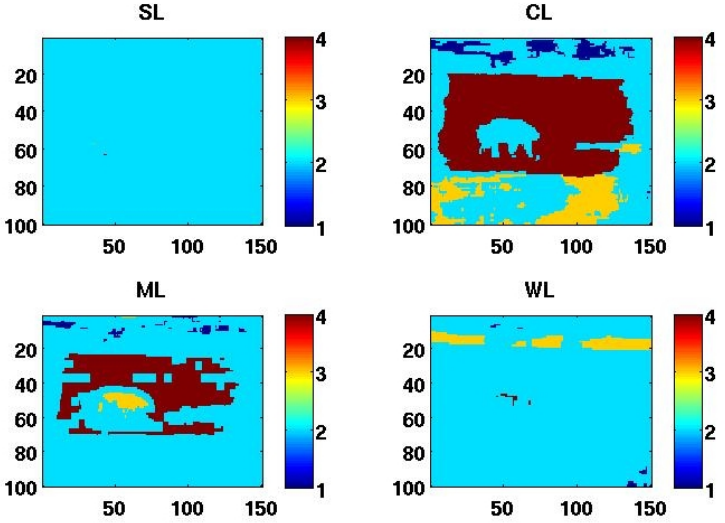
[4] R. Mall, R. Langone and J.A.K. Suykens. *Multilevel Hierarchical Kernel Spectral Clustering for Real-Life Large Scale Complex Networks*, PLOS One, e99966, 9(6), pp. 1-18, 2014.

[5] C. Alzate and J.A.K. Suykens. *Hierarchical kernel spectral clustering*, Neural Networks, 35:21-30, 2012.

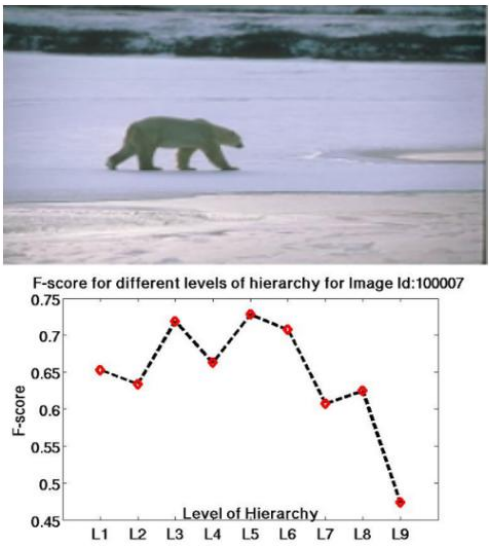
[6] A.K. Jain and P. Flynn. *Image segmentation using clustering*, In: Advances in Image Understanding, IEEE Computer Society Press, pp. 65-83, 1996.



(a) Hierarchical Segmentation by AH-KSC method from levels L1 to L9.



(b) Best results for other hierarchical techniques where F-score is maximum for $k = 4$



(c) Original Id:100007 Image & Best F-score at L5 when $k = 19$ for AH-KSC method.

Fig. 6: Hierarchical segmentation results by different methods on Id:100007 image.

[7] P. Arabie and L. Hubert. *Cluster Analysis in marketing research*, In: Advanced Methods in Marketing Research, Blackwell, Oxford, pp. 160-189, 1994.

[8] J. Hu, B.K. Ray and M. Singh. *Statistical methods for automated generation of service engagement staffing plans*, IBM J. Res. Dev. 51(3):281-293, 2007.

[9] P. Baldi and G. Hatfield. *DNA Microarrays and Gene Expression*, Cambridge University Press, 2002.

[10] A.Y. Ng, M.I. Jordan and Y. Weiss. *On spectral clustering: analysis and an algorithm*, T.G. Dietterich and S. Becker and Z. Ghahramani, editors. MIT Press: Cambridge, MA, In Proceedings of the Advances in Neural Information Processing Systems, pp. 849-856, 2002.

[11] J. Shi and J. Malik. *Normalized cuts and image segmentation*, IEEE Transactions on Pattern Analysis and Intelligence, 22(8):888-905, 2000.

[12] U. von Luxburg. *A tutorial on Spectral clustering*, Stat. Comput, 17:395-416, 2007.

[13] F.R.K. Chung. *Spectral Graph Theory*, American Mathematical Society, 1997.

[14] C. Alzate and J.A.K. Suykens. *Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(2):335-347, 2010.

[15] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle. *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.

[16] K. De Brabanter, J. De Brabanter, J.A.K. Suykens and B. De Moor. *Optimized Fixed-Size Kernel Models for Large Data Sets*, Computational Statistics & Data Analysis, 54(6):1484-1504, 2010.

[17] R. Mall, R. Langone and J.A.K. Suykens. *Self-Tuned Kernel Spectral Clustering for Large Scale Networks*, In Proceedings of the IEEE International Conference on Big Data (IEEE BigData 2013), Santa Clara, United States of America, October, pp. 385-393, 2013.

[18] M. Girolami. *Orthogonal series density estimation and the kernel eigenvalue problem*, Neural Computation, 14(3):1000-1017, 2002.

[19] R. Rabbany, M. Takaffoli and J. Fagnan and O.R. Zaiane and R.J.G.B. Campello. *Relative Validity Criteria for Community Mining Algorithms*, ASONAM, pp. 258-265, 2012.