

A Block Structure Laplacian For Hyperspectral Image Data Clustering

D. Lunga
CSIR Meraka Institute
P O Box 395
Pretoria
South Africa 0001
Email: dlunga@csir.co.za

Abstract—Over the past decade, the problem of hyperspectral data clustering has generated a growing interest from various fields including the machine learning community. This paper presents an analysis of the traditional spectral clustering approach and points to new directions that boost unsupervised pattern classification. In particular, the paper offers design insights on the generation of a well structured graph Laplacian based on an *affinity* function that induces context-dependence to create compact neighborhoods. A novel bilateral-kernel (affinity) function exploits the spatial information to generate a *diagonal-block structured* Laplacian. Experimental validations through the analysis of eigenvalues and eigenvectors demonstrate the benefits of seeking block structured affinities in hyperspectral image clustering and visualization.

I. INTRODUCTION

Hyperspectral image data clustering has received an increased attention from the statistical and machine learning communities with many applications basing their solutions on manifold learning algorithms. Such algorithms rely on the notion of nonlinear dimensionality reduction via a graph embedding framework [1]. This includes various spectral based embedding [2]–[4] and iterative gradient updates techniques [5]. In addition to clustering [6], many other benefits ranging from feature extraction [7], image segmentation [8], improved classification accuracy [1], [5], anomaly detection [9], and image visualization [10], have been proposed through graph embedding frameworks. Clustering based on a graph embedding framework hinges on the manner in which local neighborhoods are computed in the high dimensional space. A very common approach is to make use of the *heat-kernel* or Gaussian function to compute the affinity matrix that characterizes hyperspectral neighborhood graphs. The heat-kernel is known to be efficient. However, the function tends to have problems, *e.g.* on data with non-compact clusters and on data with disjoint classes as is often in many hyperspectral images. The function requires a yet to be solved problem of tuning the neighborhood width parameter in order to obtain a selection that dynamically adapts to the correlation structure of the data. Such parameters are known to be sensitive to the density of points in high dimensional space - a severe challenge that affects the stability of graph embedding algorithms. Even with a careful choice of the parameters, the computed graphs turn to have dense edge connections and do not exhibit properties

for enabling non-overlapping clusters. These challenges have a negative bearing on further analysis that maybe required on the data, *i.e.* classification, segmentation etc. This paper presents a computationally efficient *bilateral kernel* function whose properties addresses these challenges. The presentation extracts insights and further intuition by studying the impact of a spatially driven graph Laplacian. Further contributions of the study can be appreciated through its leveraging on the efficiency offered by spectral embedding and the new directions pointing towards graphical structures that boosts spectral clustering. In its assessment, the study reveals that a proper and high quality spectral decomposition of hyperspectral data begins with the induction of a block structure affinity matrix. Block structure matrices have been studied in the context of block matrices by conductivity methods [11], where instead of considering two points as similar if they are connected by a high-weight edge in the graph, an assignment of a high affinity between them is made if the overall graph conductivity between the points is high. There, graph conductivity is defined following conductivity as for electrical networks, *i.e.* the conductivity of two points depends on all paths between them. Moreover, this study reveals that for hyperspectral image data, a clear block-diagonal structure matrix can be achieved via a bilateral kernel function. A bilateral kernel function simply computes the similarity between spectral signature instances by weighting their spectral "distance" using a spatial detailed term.

Under the notion of a block-diagonal matrix, the problem of hyperspectral image clustering can now be reformulated using the similarity graph: that is, we would like to find a partition of the graph such that the edges between different groups of spectral signatures have very low weights, *i.e.* points in different clusters are dissimilar from each other, and edge connections within each spectral signature group have high weights, *i.e.* points within the same cluster are similar to each other. While a robust theoretical approach to motivate the choice of an affinity function is of relevance and great interest to this work, it is not the main focus of this presentation. However, an attempt is made to present strong insights and motivation through an empirical guide that analyzes and compares properties of three affinity functions that represents local spectral signature neighborhoods. Even though

our mention of clustering based on various dimensionality reduction techniques remains relevant, care is taken not to deluge the presentation with details of such methods but rather analyze the computed graph Laplacians through a spectral clustering approach which was chosen solely for its efficiency.

The paper is structured as follows. A description of hyperspectral neighborhood graphs based on the heat-kernel, the normalized heat-kernel and the bilateral kernel functions is introduced in Section II. Two general approaches on the formulation of graph Laplacians and their properties are presented in Section III. In Section IV, a description of the data and experimental analysis on the eigenvalues, eigenvectors, block-diagonal structures and hyperspectral clusters is presented. Finally, Section V concludes with a summary and future work ideas.

II. HYPERSPECTRAL NEIGHBORHOOD GRAPHS

Given an image dataset with training samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ in \mathbb{R}^m (m -dimensional spectral feature space) and n is the total number of training samples, data clustering via dimensionality reduction algorithms adapts a graph embedding framework in which $G = \{\mathbf{X}, \mathbf{W}\}$ is the undirected weighted graph and \mathbf{W} is the $n \times n$ data dependent similarity or *affinity* matrix. The algorithms utilize the notion of affinity weights $\mathbf{W}_{ij} \in [0, 1]$ to measure the "distance" between two sample observations. The affinity functions do not utilize class label information, but rather characterizes the neighborhood relationships between all pairs of points based on feature differences. The characterization of relations in a given dataset can be pursued in various forms. Determining a suitable affinity function and eventually the neighborhood graph often involves complex strategies for tuning parameters that include for example - setting the width of the neighborhood regions. As to be illustrated in this paper, neighborhood graphs that benefits from spectral clustering algorithms tend to conform to a particular structure, *i.e.* the block-diagonal structure. This is a very important insight not only in remote sensing data but also in other related fields where research efforts in graph embedding methods continue to emphasize the fine tuning of various parameters and proposal of nonlinear objective functions that require complex optimization strategies.

Neighborhood construction for graph embedding starts with the choice of the affinity function \mathbf{W}_{ij} . The commonly adapted frameworks try to ensure that the local neighborhoods induced by the affinity function extract coherent and "meaningful" structures. In general, there is no consensus to the definition of "meaningful" structures. This study adapts on its reference to groupings of similar observed spectral signatures as measured by the reflectance values of each image pixel.

A. Heat-Kernel Based Graph

A very popular approach to measure the affinity between hyperspectral data samples \mathbf{x}_i and \mathbf{x}_j is the *heat-kernel* function,

$$\mathbf{W}_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\tau_i \tau_j}\right), \quad (1)$$

where $\tau_i = \|\mathbf{x}_i - \mathbf{x}_i^{(k_{nn})}\|$ denotes the local scaling of data samples in the neighborhood of \mathbf{x}_i , and $\mathbf{x}_i^{(k_{nn})}$ is the k_{nn} -nearest neighbor of \mathbf{x}_i . Without a threshold set on the affinity values, construction of graph based on this function results in dense and in most cases yields a single completely connected graph component. Although it has been shown to result in effective locality preserving properties for graph embedding of various datasets, further improvements towards sparse affinity neighborhoods can be achieved by adapting the scaling parameter τ_i to the local data statistics which often provide a stronger adaptivity to the underlying structure of the embedded image manifolds.

B. Normalized Heat-Kernel Based Graph

In a stochastic neighbor embedding [12] framework or technique for preserving probabilities on lower dimensional coordinate systems that are nonlinear, a normalized variant of the heat-kernel was proposed. There, the notion of affinities or similarities was interpreted as the probability of neighboring instances choosing one another as neighbors. Such probabilities are computed from the normalized Gaussian(or heat-kernel) functions

$$\mathbf{W}_{ij} = \frac{\exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_i}\right\}}{\sum_{r \neq i} \exp\left\{-\frac{\|\mathbf{x}_r - \mathbf{x}_i\|^2}{2\sigma_i}\right\}} \quad (2)$$

The significance of a normalized affinity value is simply due to the strategy used to choose the parameter σ_i . It is computed using a binary search method ensuring that the entropy of the distribution \mathbf{W}_i is approximately $\log(k)$, with k being the effective number of neighbors. As such, the density of the data samples is adapted to inform the nature of neighborhood regions - a very crucial element in determining how far a neighborhood should stretch in a high dimensional space.

C. Bilateral Kernel Based Graph

Furthermore, it is well known that spatial preprocessing methods are often applied to remove noise and smooth images. In other applications they can also enhance spatial texture information resulting in features that improve the performance of classification techniques. For example in [13], nonlinear diffusion partial differential equations (PDEs) and wavelet shrinkage were used for spatial preprocessing of hyperspectral images, and the results obtained demonstrated a significant improvement on classification performance. Similarly, neighborhood affinities can also be defined by introducing context in the form of spatial weighting of spectral values. This can be accomplished by computing the "distance" between two hyperspectral pixels in a separable approach to introduce spatially sensitive affinities through a bilateral kernel function $\mathbf{W}_{ij} = W(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}_i, \mathbf{x}_j)$ as

$$W(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}_i, \mathbf{x}_j) = \exp\left\{\frac{-\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\sigma_s^2}\right\} \cdot \tilde{W}_p(\mathbf{x}_i, \mathbf{x}_j) \quad (3)$$

where \mathbf{s}_i denotes the spatial coordinates of pixel i , \mathbf{x}_i denotes the photometric m -dimensional spectral vector. The expression

$\|s_i - s_j\|^2$ weights image pixel values as a function of the spatial distance from the center pixel and σ_s is the variance parameter. The kernel also employs a nonlinear term:

$$\tilde{W}_p(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \mathbf{x}_j)^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j) \right\} \quad (4)$$

which simply weights relations as a function of spectral differences between the center pixel and its neighbor pixel. With additional manipulations as shown in [10], $\tilde{W}_p(\mathbf{x}_i, \mathbf{x}_j)$ can be rewritten as

$$\tilde{W}_p(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ \frac{-N}{2} \text{tr}(\Sigma^{-1} \mathbf{S}) \right\} \quad (5)$$

where \mathbf{S} is the sample covariance, $\mathbf{S} = \frac{1}{N} \mathbf{X} \mathbf{X}^T$. To compute $\tilde{W}_p(\mathbf{x}_i, \mathbf{x}_j)$, decompose the true covariance matrix into a product $\Sigma = \mathbf{E} \Lambda \mathbf{E}^T$, where \mathbf{E} is the orthogonal eigenvector matrix and Λ is the corresponding diagonal matrix of eigenvalues. Adapting the efficient sparse matrix transform (SMT) approach in estimating the covariance matrix Σ [14], the optimization problem can be solved as $\hat{\mathbf{E}} = \text{argmin}_{\mathbf{E} \in \Omega} \{ |\text{diag}(\mathbf{E}^T \mathbf{S} \mathbf{E})| \}$, and the optimal eigenvalue matrix is obtained from $\hat{\Lambda} = \text{diag}(\hat{\mathbf{E}}^T \mathbf{S} \hat{\mathbf{E}})$, where Ω is the set of allowed orthogonal transforms that can be computed using a series of *Givens rotations* [14]. A simple manipulation can show that $\Sigma^{-1} = \hat{\mathbf{E}} \hat{\Lambda}^{-1} \hat{\mathbf{E}}^T$. This approach to computing the covariance matrix Σ is efficient and robust in handling the singularities of Σ in the high dimensional space of hyperspectral image bands.

III. GRAPH LAPLACIANS

In general, given neighborhood graph G , the most challenging issue is the unsupervised classification of measured patterns into groups according to their similarity. To this note the graph Laplacian \mathbf{L} plays an important role in such problems as it provides a link for projection of data onto new coordinate system via graph embedding algorithms. Its properties, through eigenvalues and eigenvectors information, can be used to find the number of spanning trees as well as approximating sparse graph cuts that lead to meaningful clusters. Two approaches for computing the graph Laplacian are discussed next.

A. Unnormalized Graph Laplacian

The unnormalized Laplacian matrix is simply computed from a difference between the degree matrix \mathbf{D} , with $D_{ii} = \sum_j \mathbf{W}_{ij}, \forall i$ and the affinity matrix \mathbf{W} . That is $\mathbf{L} = \mathbf{D} - \mathbf{W}$. The few properties that are of interest for unsupervised pattern classification are summarized as follows.

Proposition 1. (*Laplacian Properties*). *The matrix \mathbf{L} satisfies the following properties:*

- For every vector $\mathbf{h} \in \mathbb{R}^p$ we have

$$\mathbf{h}^t \mathbf{L} \mathbf{h} = \frac{1}{2} \sum_{i,j=1}^p \mathbf{W}_{ij} (\mathbf{h}_i - \mathbf{h}_j)^2$$

- \mathbf{L} is symmetric and positive semi-definite.

- The smallest eigenvalue of \mathbf{L} is 0, whose corresponding eigenvector is a constant vector $\mathbf{1}$.
- \mathbf{L} has p non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$.

Additional properties of unnormalized and normalized Laplacian matrices including proofs for the above proposition can be found in [6], [15]–[17]. Of more importance for data clustering applications is the following:

Proposition 2. (*The number of connected components and spectrum of \mathbf{L}*). *Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of \mathbf{L} equals the number of connected components C_1, C_2, \dots, C_k in the graph. The eigenspace of the eigenvalue 0 is spanned by the indicator vectors $\mathbf{1}_{C_1}, \dots, \mathbf{1}_{C_k}$ of those components.*

Proof. Case I $k = 1$. That is consider a graph that is completely connected. Assume that \mathbf{h} is an eigenvector with eigenvalue 0. Then we know that

$$0 = \mathbf{h}^t \mathbf{L} \mathbf{h} = \sum_{i,j=1}^p \mathbf{W}_{ij} (\mathbf{h}_i - \mathbf{h}_j)^2 \quad (6)$$

Since each $\mathbf{W}_{ij} \geq 0$, $\mathbf{h}^t \mathbf{L} \mathbf{h}$ can only equal zero if all the terms $\mathbf{W}_{ij} (\mathbf{h}_i - \mathbf{h}_j)^2$ in the summation are equal to zero. This means that for two hyperspectral pixels with vertices denoted by \mathbf{x}_i and \mathbf{x}_j that are connected, *i.e.* $\mathbf{W}_{ij} > 0$, their corresponding functions \mathbf{h}_i and \mathbf{h}_j have to be equal for (6) to vanish. With this intuition, it is simple to deduce that \mathbf{h} has to be more or less constant for all vertices that can be connected by a path in the graph. For a graph with one giant connected component we thus only have the constant vector $\mathbf{1}$ with corresponding eigenvalue 0.

Case II consider a case of $k > 1$ connected components. Without loss of generality assume that the vertices are ordered according to the connected components they belong to. In this case, the affinity matrix \mathbf{W} has a block diagonal form, and the same is true for the Laplacian matrix \mathbf{L} :

$$\mathbf{L} = \mathbf{L}_1 \oplus \mathbf{L}_2 \oplus \dots \oplus \mathbf{L}_k = \begin{pmatrix} \mathbf{L}_1 & & & \\ & \mathbf{L}_2 & & \\ & & \ddots & \\ & & & \mathbf{L}_k \end{pmatrix}$$

where \oplus denotes a direct sum of matrices. Each of the blocks \mathbf{L}_i is a well defined graph Laplacian matrix corresponding to the subgraph of the i th connected component. The full spectrum of \mathbf{L} is obtained from the spectra union of all \mathbf{L}_i block matrices. The eigenvectors of \mathbf{L} are the eigenvectors of all \mathbf{L}_i blocks combined. Furthermore, since each \mathbf{L}_i relates to a connected component, from *Proposition 1*, we know that every \mathbf{L}_i has eigenvalue 0 with multiplicity 1, and its corresponding eigenvector is the constant $\mathbf{1}$ vector on the i th connected component. In theory, the matrix \mathbf{L} has as many zero eigenvalues as there are connected components, and the corresponding eigenvectors are the indicator vectors of the connected components. \square

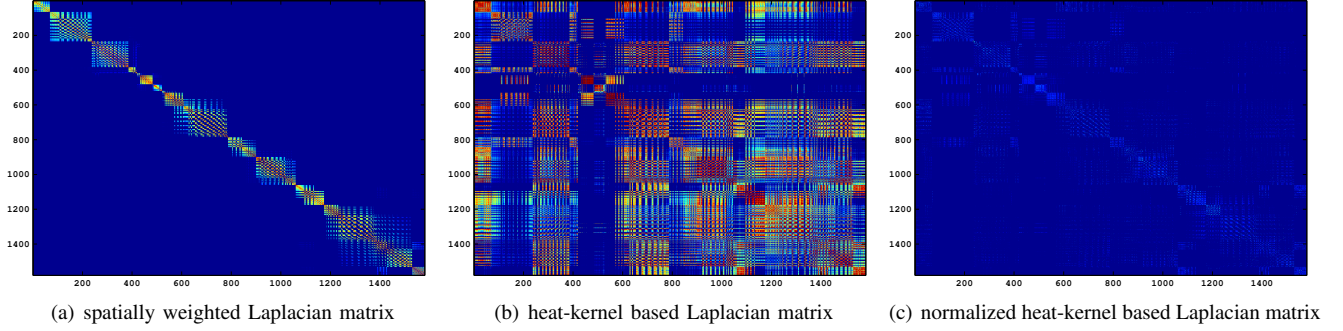


Fig. 3. Botswana data Laplacian matrices. Pure blue color indicates small affinity values while red-yellow colors indicates neighborhoods with high weights in the graph.

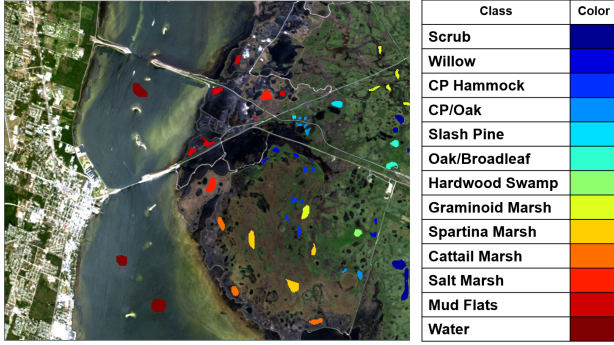


Fig. 1. Colored ground reference of AVIRIS hyperspectral image over Kennedy Space Center (KSC), Florida.

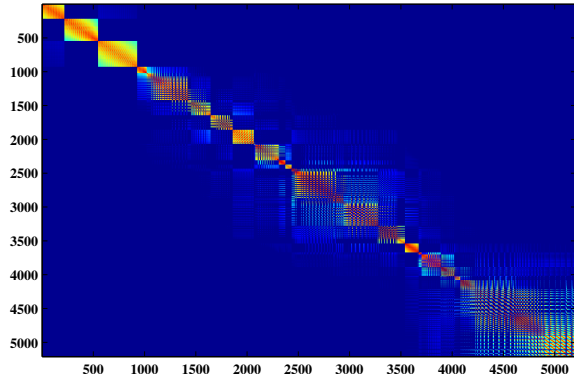


Fig. 2. A spatially weighted Laplacian matrix for KSC data. Pure blue color indicates small affinity values while red-yellow colors indicates neighborhoods with high edge weights.

B. Normalized Graph Laplacian

A normalized graph Laplacian is commonly studied under the following two forms:

$$\begin{aligned} \mathbf{L}_n &= \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \\ \mathbf{L}_{rw} &= \mathbf{I} - \mathbf{D}^{-1} \mathbf{W} \end{aligned}$$

where \mathbf{L}_n denotes the symmetric matrix and \mathbf{L}_{rw} relates to the notion of random walks.

Proposition 3. (Normalized Laplacian properties) Both ver-

sions of the normalized graph Laplacian matrices satisfy the following,

- For every $\mathbf{h} \in \mathbb{R}^p$, we have

$$\mathbf{h}^t \mathbf{L}_n \mathbf{h} = \frac{1}{2} \sum_{i,j=1}^p \mathbf{W}_{ij} \left(\frac{\mathbf{h}_i}{\sqrt{d_i}} - \frac{\mathbf{h}_j}{\sqrt{d_j}} \right)^2$$

- λ is an eigenvalue of \mathbf{L}_{rw} with eigenvector u iff λ is an eigenvalue of \mathbf{L}_n with eigenvector $w = \mathbf{D}^{1/2} u$.
- λ is an eigenvalue of \mathbf{L}_{rw} with eigenvector u iff λ and u solve the generalized eigen-problem $\mathbf{L} u = \lambda \mathbf{D} u$.
- 0 is an eigenvalue of \mathbf{L}_{rw} with the constant $\mathbf{1}$ vector. 0 is an eigenvalue of \mathbf{L} with eigenvector $\mathbf{D}^{1/2} \mathbf{1}$.
- \mathbf{L}_n and \mathbf{L}_{rw} are positive semi-definite and have p non-negative real-valued eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_p$

Similarly as in the case of unnormalized Laplacian, the multiplicity of the eigenvalue 0 is related to the number of connected components. This can be proved in a similar fashion to Proposition 2.

IV. EXPERIMENTS

To illustrate the effect of seeking a block diagonal graph Laplacian for hyperspectral neighborhood graphs, we consider two separate data sets. First a hyperion data with nine identified classes of complex natural vegetation acquired over the Okavango Delta, Botswana, in May 2001, [18]. The general class groupings include seasonal swamps, occasional swamps, and woodlands. Signatures of several classes are spectrally overlapped, typically resulting in poor classification accuracies. After removing water absorption, noisy, and overlapping spectral bands, 145 bands were used for spectral embedding.

A second dataset, shown in Figure 1, consist of a hyperspectral data that was acquired by the National Aeronautics and Space Administration (NASA) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor at 18-m spatial resolution over Kennedy Space Center during March 1996. Noisy and water absorption bands were removed, leaving 176 features for thirteen wetland and upland classes of interest. Cabbage Palm Hammock and Broad Leaf/Oak Hammock are upland trees; Willow Swamp, Hardwood Swamp, Graminoid Marsh and Spartina Marsh are trees and grasses in wetlands.

Their spectral signatures are mixed and often exhibit only subtle differences posing some challenges for clustering algorithms.

A. Hyperspectral Laplacian Eigenspectrum Analysis

Figure 3 shows a result from incorporating spatially weighting. The plot demonstrates that a bilateral kernel function infuses local adaptivity and spatial sensitivity in affinity computations. Such an adaptation leads to a block-diagonal structure graph Laplacian and ultimately enables preservation of local disjoint neighborhoods that are compact and similar. Further analysis of the Laplacian matrices, following Proposition 2, shows that its resulting eigenspectra and corresponding eigenvectors do contain useful details that can be used in unsupervised pattern classification.

1) *Eigenvalues*: The eigenvalues of the graph Laplacian can be used to identify the number of cluster components. A widely used eigengap heuristic [6], [17] can be adapted. The goal is to choose the number k such that all eigenvalues $\lambda_1, \dots, \lambda_k$ are very small, but λ_{k+1} is relatively large. The motivation bears on intuition from the perturbation theory, where an observation can be made in the ideal case of k completely disconnected clusters, the eigenvalue 0 has multiplicity k , and then there is a gap to the $(k+1)$ th eigenvalue, where $\lambda_{k+1} \geq 0$. For data with well pronounced clusters the heuristic is very effective. For example, by zooming-in on the plots in Figure 4, there are twenty identifiable eigengaps (or graph cuts/clusters) on the eigenvalues plot obtained with the bilateral kernel based Laplacian. Only two graph cuts could be identified from the unnormalized heat-kernel based graph Laplacian and no cuts were observed on the heat-kernel based Laplacian. This result demonstrates the superiority of the proposed bilateral kernel function in capturing distinct separations for different spectral signatures.

2) *Eigenvectors*: Given a block structure Laplacian, as computed by the bilateral kernel, data clustering can be performed by observing that entry values within selected eigenvectors corresponding to small valued eigenvalues tend to have a disjoint density of points that groups similar spectral signature patterns. This is shown in Figure 5(a) for the Botswana data set. A similar result, even though with diminished quality, can be observed from the eigenvectors computed by the normalized heat kernel. Following the result in Figure 5 and the analysis of eigenvectors in general, we can deduce a very simple and important insight, *i.e.* a block-diagonal structure Laplacian matrix generates corresponding eigenvectors that splits the data into a hierarchy of subclusters corresponding to varying densities in the observation samples. Such a characteristic could further be exploited for further studies in data segmentation as well as object classification applications. Figure 6 illustrates on this further by showing a 2-dimensional principal component analysis projection of the Botswana image data that is color-coded by the entry values of the 3rd and 4th eigenvectors presented in Figure 5. A similar result that corresponds to computed KSC eigenvectors is presented in Figure 7. From both figures, it is clear that even

without the ground truth labels, one can simply make use of the eigenvector values computed from the spatially sensitive Laplacian, to obtain the graph cuts that correspond to different spectral signature classes.

B. Hyperspectral Image Data Clustering

Most image data clustering applications seek to project data onto a coordinate system spanned by smallest eigenvectors of the graph Laplacian. In this section, we illustrate empirically that obtaining a block structure Laplacian ultimately enables preserving local disjoint neighborhoods that are compact and as such may benefit most clustering algorithms that are based on the graph embedding framework. Spectral clustering is one such algorithm that is widely applied. The method has a strong bearing on the structure of the graph Laplacian. In Figure 4, the first few eigenvalues of both the spatially weighted and the normalized kernel based Laplacians are zero, and they have a corresponding number of cluster indicator eigenvectors that are constant. The reason for this is simple, *i.e.* due to the diagonal-block structure data gets clustered in correspondence with disconnected neighborhood subgraphs that are associated with the indicator eigenvectors. Contrary to the widely applied notion that the first k eigenvectors from a spectral embedding approach often provide the best clustering of data, Figure 8 shows a result that omits the 1st eigenvector (which is obviously constant from Proposition 1), 2nd and 3rd eigenvectors (neglected for lack of detail to establish different clusters) and a projection of data onto pairwise dimensions spanned by combinations of 4th, 5th, 6th and 7th eigenvectors. The coordinates based on these eigenvectors reveal a visualization that identifies meaningful spectral signature separations. As in most post dimensionality reduction processing, a K-means algorithm can be applied on the projected coordinates to further determine cluster means that may be useful in other applications.

Due to the difficulty in providing quantitative analysis for clustering algorithms, progress in fully appreciating spectral based algorithms lags as the task of clustering in general terms is yet to be completely defined. However as shown in various studies [3], [5], [19]–[21], where ground truth exists, classification algorithms can be designed to take advantage of the embedding obtained with a bilateral kernel function.

V. CONCLUSIONS AND FUTURE WORK

Clustering details for hyperspectral image data often lie on sparse, nonlinear coordinate system whose geometric and topological structures can be exploited with graph embedding framework. This study concentrated on demonstrating new insights and a methodology for achieving high quality clustering results with hyperspectral data. Through experimental demonstrations, the paper suggests a particular construction of affinity matrices that achieve a block structure diagonal Laplacian. A diagonal block affinity matrix is shown to boost spectral embedding with benefits for unsupervised pattern classification applications. With the introduction of spatial details, an automatic induction of sparsity neighborhoods

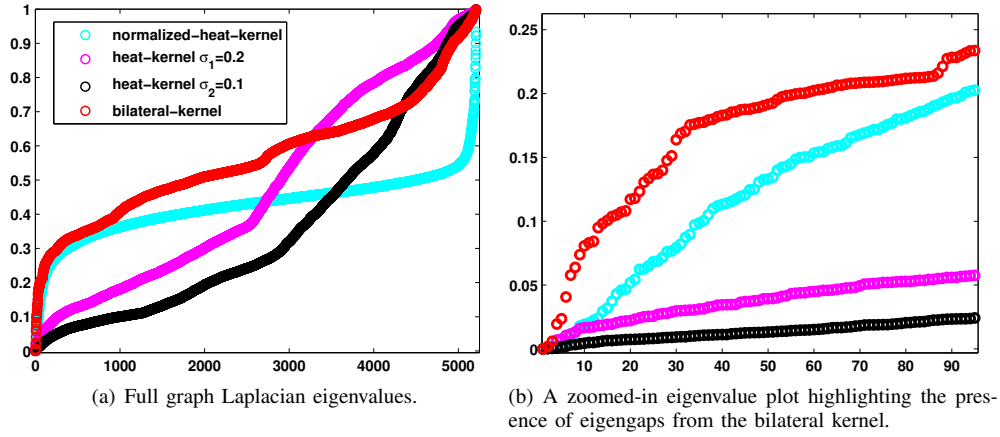


Fig. 4. Laplacian eigenvalues for the KSC data set.

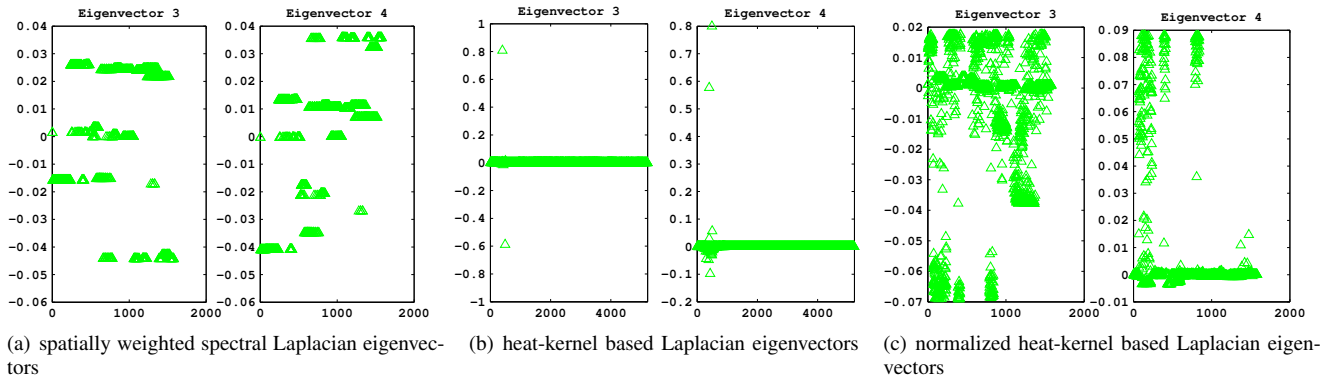


Fig. 5. Botswana data Laplacian eigenvectors.

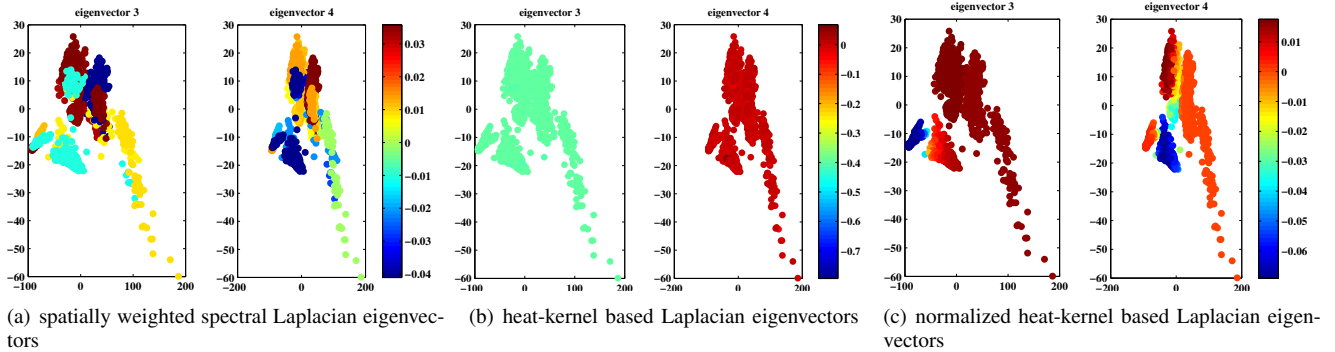


Fig. 6. 2 dimensional PCA projected Botswana data color-coded by the 3rd and 4th eigenvector values of different Laplacian matrices.

with a natural ordering on the rows of the Laplacian matrix is achieved. Spectral data clustering is pursued through an analysis of the Laplacian eigenspectrum and its corresponding eigenvectors, with various levels of class separation displayed for different paired eigenvectors.

VI. ACKNOWLEDGEMENTS

The author would like to thank Dr Melba Crawford, Purdue University, and Dr Saurabh Prasad, University of Houston, for the helpful discussions on the content of the paper. I would

like to also thank the anonymous reviewers who pointed out improvements for the final presentation of the paper.

REFERENCES

- [1] M. M. Crawford, L. Ma, and W. Kim, "Exploring nonlinear manifold learning for classification of hyperspectral data," in *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, S. Prasad, J. Chanussot, and L. B. (Eds), Eds. London: Springer Verlag, 2011.
- [2] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185–1198, 2012.

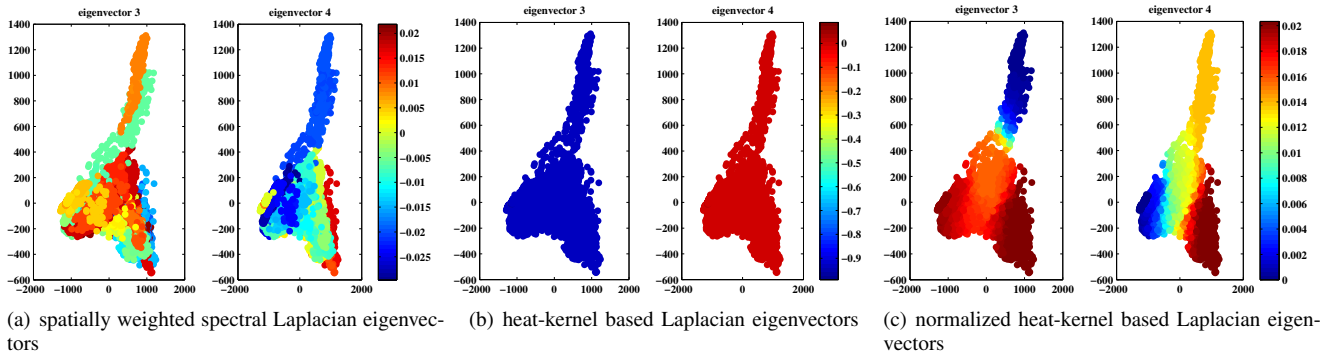


Fig. 7. 2 dimensional PCA projected KSC data color-coded by the 3rd and 4th eigenvector values of different Laplacian matrices.

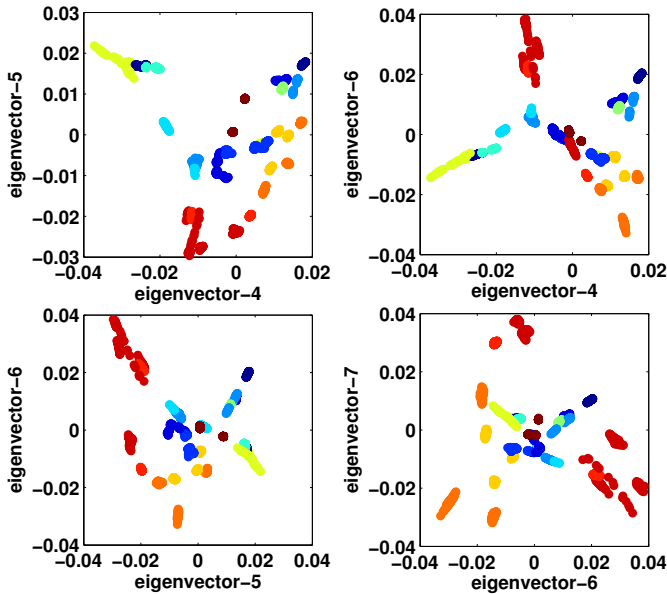


Fig. 8. 2D scatter plots for various dimensions (or eigenvectors) of spectral embedding of 5211 KSC data samples.

- [3] Y.-L. Chang, J.-N. Liu, C.-C. Han, and Y.-N. Chen, "Hyperspectral image classification using nearest feature line embedding approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–10, 2013.
- [4] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 895–898, 2011.
- [5] D. Lunga and O. Ersoy, "Multidimensional artificial field embedding with spatial sensitivity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–1, 2013.
- [6] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, December 2007. [Online]. Available: <http://dx.doi.org/10.1007/s11222-007-9033-z>
- [7] J. He, L. Zhang, Q. Wang, and Z. Li, "Using diffusion geometric coordinates for hyperspectral imagery representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 767–771, 2009.
- [8] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 206–210, 2007.
- [9] L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, 2013.
- [10] D. Lunga and O. Ersoy, "Spherical stochastic neighbor embedding of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 857–871, 2013.
- [11] I. Fischer, "Amplifying the block matrix structure for spectral clustering," in *In IDSIA*, 2005, pp. 21–28.
- [12] G. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proc. of ICML*, vol. 15, 2002, pp. 833–840.
- [13] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE TGRS*, vol. 6, no. 2, pp. 297–301, 2009.
- [14] J. Theiler, G. Cao, L. R. Bachega, and C. A. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 424–437, 2011.
- [15] B. Mohar, "The laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*. Wiley, 1991, pp. 871–898.
- [16] —, "Some applications of laplace eigenvalues of graphs," in *GRAPH SYMMETRY: ALGEBRAIC METHODS AND APPLICATIONS, VOLUME 497 OF NATO ASI SERIES C*. Kluwer, 1997, pp. 227–275.
- [17] F. R. K. Chung, *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, 1997.
- [18] A. L. Neuenschwander, *Remote sensing of vegetation dynamics in response to flooding and fire in the Okavango Delta-Botswana*. Univ. Texas Austin, Austin, TX: Ph.D. dissertation, 2007.
- [19] B. Hou, X. Zhang, Q. Ye, and Y. Zheng, "A novel method for hyperspectral image classification based on laplacian eigenmap pixels distribution-flow," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1602–1618, 2013.
- [20] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.
- [21] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, March 2013.