

HYPERSPECTRAL BAND SELECTION BASED ON MATRIX CUR DECOMPOSITION

Katherine Henneberger¹, Longxiu Huang² and Jing Qin¹

¹ Department Mathematics, University of Kentucky, Lexington, KY 40506, USA

² Department of of CMSE and Mathematics, Michigan State University, East Lansing, MI 48824, USA.

ABSTRACT

Band selection is an important technique for eliminating spectral redundancy of hyperspectral imagery (HSI) while preserving critical information. Recently, correlations among neighboring bands or pixels have been exploited in the form of graph regularizations to reduce the data dimensionality efficiently. However, manipulation of graph regularizations typically causes computational bottlenecks. In this work, we propose a robust method for hyperspectral band selection based on spatial/spectral graph Laplacians and matrix CUR decomposition. The efficiency of the proposed method has been shown on two real data sets by comparing with several other state-of-the-art band selection methods.

Index Terms— Hyperspectral band selection, matrix CUR decomposition, classification, robust PCA

1. INTRODUCTION

Hyperspectral imagery (HSI) is an advanced technology that gathers spectral data from ground objects in the form of hundreds of narrow bands, which can provide more spectral information than classical RGB imagery and has been successfully applied in various fields such as in remote sensing, biomedicine, agriculture, and art conservation. However, the high dimensionality of HSI poses challenges such as spectral redundancy and computational burden. Therefore, it is necessary to consider dimensionality reduction and redundancy removal for HSI data. Hyperspectral band selection aims to alleviate these challenges by selecting a subset of original bands while preserving important spectral information [1].

In general, HSI band selection methods can be classified into supervised and unsupervised methods. Supervised band selection requires prior knowledge of the HSI data. However, training samples with annotated class labels are often either limited or difficult to obtain. Therefore, unsupervised methods are preferable in many applications [1]. Some state-of-the-art unsupervised HSI band selection algorithms include enhanced fast density-peak-based clustering (E-FDPC) algorithm [2], fast neighborhood grouping (FNGBS) method [3],

and similarity-based ranking strategy with structural similarity (SR-SSIM) [4].

Recently, a fast and robust principal component analysis on Laplacian graph (FRPCALG) method for band selection is proposed in [5] by considering the band-wise graph structure. Specifically, it incorporates a graph regularization based on the band-wise graph Laplacian into the robust PCA framework so that spectral correlation is preserved in a sparse and low-rank data matrix. However, solving the nuclear norm regularization typically requires the expensive singular value decomposition (SVD). In addition, it fails to consider spatial correlation of the data matrix. To address these issues, we propose a novel hyperspectral band selection method based on the spatial and spectral graph regularizations, and also introduce the CUR decomposition to handle the low-rank constraint in the alternating direction method of multipliers (ADMM) algorithm framework. Numerical results on two real data sets have shown the great potential of our method.

The rest of the paper is organized as follows: Section 2 provides some preliminaries that will be used throughout this work. Section 3 describes the proposed band selection method in detail. Various numerical experiments on geometrically deformed remote sensing images are shown in Section 4. Finally, the conclusion and some future works are summarized in Section 5.

2. PRELIMINARIES

In this section, we briefly introduce the construction of spatial and spectral graphs and the matrix CUR decomposition.

The graph Laplacian plays an important role in band selection to preserve local structures across the spectral bands. Consider a hyperspectral data matrix $B \in \mathbb{R}^{\ell \times n}$ where ℓ is the number of pixels and n is the number of bands. The rows and columns correspond to the spatial and spectral samples respectively. In the spectral domain, consider a graph $G_c := (\mathcal{V}_c, \mathcal{E}_c)$ where $\mathcal{V}_c = \{b_i^c\}_{i=1}^n$ is the set of vertices (columns of B) and $\mathcal{E}_c$ is the set of edges. Let $W_c = \{(W_c)_{ij}\} \in \mathbb{R}^{n \times n}$ be the weighted matrix where each component is computed such that $(W_c)_{ij}$ is a positive value proportional to the degree of similarity of pixels i and j , but zero when i and j are dissimilar [6]. We consider the k -nearest neighbors to determine

The research of K. Henneberger and J. Qin is supported by the National Science Foundation Grant DMS 1941197. L. Huang was partially supported by the AMS Simons Travel Grant.

proximity and utilize the heat kernel,

$$(W_c)_{ij} = \begin{cases} \exp\left(-\frac{\|b_i^c - b_j^c\|_2^2}{\sigma_c^2}\right), & \text{if } b_i^c \text{ and } b_j^c \text{ are neighbors;} \\ 0, & \text{otherwise,} \end{cases}$$

where $\sigma_c > 0$ is the kernel parameter. Next we define the symmetrically normalized spectral graph Laplacian $\Phi_c \in \mathbb{R}^{n \times n}$ given by $\Phi_c = I_n - D_c^{-1/2} W_c D_c^{-1/2}$ where I_n is the $n \times n$ identity matrix, and D_c is the diagonal degree matrix defined as $(D_c)_{ij} = \sum_{j=1}^n (W_c)_{ij}$.

Similarly, we consider the spatial neighbor graph $\mathbf{G}_s := (\mathcal{V}_s, \mathcal{E}_s)$ where $\mathcal{V}_s = \{b_i^s\}_{i=1}^\ell$ is the set of vertices (rows of B) and $\mathcal{E}_s$ is the set of edges. In contrast to the spectral case, we consider a different construction of the weighted matrix $W_s = \{(W_s)_{ij}\}$. In the spatial domain, we consider the patchwise similarity [7]. In particular, we compute

$$(W_s)_{ij} = \begin{cases} \exp\left(-\frac{\|\mathcal{N}(b_i^s) - \mathcal{N}(b_j^s)\|_2^2}{\sigma_s^2}\right), & \text{if } b_i^s \text{ and } b_j^s \\ & \text{are neighbors;} \\ 0, & \text{otherwise,} \end{cases}$$

where σ_s is the kernel parameter, and $\mathcal{N}(b_i^s) \in \mathbb{R}^{p^2 \times n}$ is a reshaped patch of size $p \times p$ centered at the i -th pixel. The symmetrically normalized graph Laplacian in the spatial domain is defined as $\Phi_s = I_\ell - D_s^{-1/2} W_s D_s^{-1/2}$ where the diagonal matrix D_s has i -th diagonal entry $(D_s)_{ii} = \sum_{j=1}^\ell (W_s)_{ij}$.

Furthermore, CUR decompositions are efficient low-rank matrix approximation methods that use a small number of actual columns and rows of the original matrix. In this paper, we adopt the following definition.

Definition 2.1 ([8]) Given $Y \in \mathbb{R}^{\ell \times n}$, let $C \in \mathbb{R}^{\ell \times s_c}$ be a column submatrix of Y with column size s_c , $R \in \mathbb{R}^{s_r \times n}$ a row submatrix of Y with row size s_r , and $U \in \mathbb{R}^{s_r \times s_c}$ the intersection part of C and R . The CUR decomposition of Y is $\hat{Y} = CU^\dagger R$, where $U^\dagger$ is the pseudoinverse of U .

3. PROPOSED ALGORITHM

Given a hyperspectral data matrix $Y \in \mathbb{R}^{\ell \times n}$ with ℓ spatial pixels and n spectral bands, we first use Y to construct its spatial and spectral graph Laplacians as in Section 2. To select bands, we aim to decompose Y into a sum of low-rank matrix B and a sparse matrix S via the following model

$$\min_{\text{rank}(B) \leq r, S} \|Y - B - S\|_1 + \lambda \|S\|_1 + \frac{\gamma_1}{2} \text{tr}(B^T \Phi_s B) + \frac{\gamma_2}{2} \text{tr}(B \Phi_c B^T) \quad (1)$$

where $\Phi_s \in \mathbb{R}^{\ell \times \ell}$ is the graph Laplacian in the spatial domain and $\Phi_c \in \mathbb{R}^{n \times n}$ is the graph Laplacian in the spectral domain. In order to apply the ADMM framework to minimize

(1), we introduce the auxiliary variable Z and rewrite (1) into an equivalent form as follows

$$\min_{\text{rank}(B) \leq r, Z, S} \|Z\|_1 + \lambda \|S\|_1 + \frac{\gamma_1}{2} \text{tr}(B^T \Phi_s B) + \frac{\gamma_2}{2} \text{tr}(B \Phi_c B^T) \\ \text{s.t. } Z = Y - B - S.$$

We introduce an indicator function to take care of the nonlinear constraint about B . Let $\Pi = \{X \in \mathbb{R}^{\ell \times n} | \text{rank}(X) \leq r\}$. The indicator function χ_Π is defined as $\chi_\Pi(X) = 0$ if $X \in \Pi$ and ∞ otherwise. Then the augmented Lagrangian can be written as

$$\mathcal{L} = \|Z\|_1 + \lambda \|S\|_1 + \frac{\gamma_1}{2} \text{tr}(B^T \Phi_s B) + \frac{\gamma_2}{2} \text{tr}(B \Phi_c B^T) \\ + \chi_\Pi(B) + \frac{\beta}{2} \|Y - B - S - Z + \tilde{Z}\|_F^2$$

where $\tilde{Z}$ is a dual variable and $\beta > 0$ is the penalty parameter. The resulting algorithm can be described as follows:

$$\begin{cases} B \leftarrow \underset{\text{rank}(B) \leq r}{\text{argmin}} \frac{\beta}{2} \|Y - B - S - Z + \tilde{Z}\|_F^2 \\ \quad + \frac{\gamma_1}{2} \text{tr}(B^T \Phi_s B) + \frac{\gamma_2}{2} \text{tr}(B \Phi_c B^T) \\ S \leftarrow \underset{S}{\text{argmin}} \lambda \|S\|_1 + \frac{\beta}{2} \|Y - B - S - Z + \tilde{Z}\|_F^2 \\ Z \leftarrow \underset{Z}{\text{argmin}} \|Z\|_1 + \frac{\beta}{2} \|Y - B - S - Z + \tilde{Z}\|_F^2 \\ \tilde{Z} \leftarrow \tilde{Z} + Y - B - S - Z \end{cases}$$

Then the ADMM algorithm requires solving three subproblems at each iteration. Specifically, we aim to minimize $\mathcal{L}$ with respect to B , S , and Z . A common approach for updating B is to use the skinny SVD, however this can be costly when the size of our matrix is large. Instead we utilize the CUR decomposition. The variable B can be updated as

$$B^{j+1} = \underset{\text{rank}(B) \leq r}{\text{argmin}} \frac{\beta}{2} \|Y - B - S^j - Z^j + \tilde{Z}\|_F^2 \\ + \frac{\gamma_1}{2} \text{tr}(B^T \Phi_s B) + \frac{\gamma_2}{2} \text{tr}(B \Phi_c B^T) := f(B)$$

Then utilizing gradient descent with step size τ , we update $B^{j+1} = B^j - \tau \nabla f(B^j)$ where the gradient is calculated as $\nabla f(B^j) = \beta(B^j - (Y - S^j - Z^j + \tilde{Z})) + \gamma_1 \Phi_s B^j + \gamma_2 B^j \Phi_c$. The CUR decomposition is updated as

$$C \leftarrow C - \tau \nabla f(B^j)(:, J) \\ R \leftarrow C - \tau \nabla f(B^j)(I, :) \\ U \leftarrow \frac{1}{2}(C + R)$$

where I and J are the respective row and column index sets. Then

$$B^{j+1} = CU^\dagger R. \quad (2)$$

Holding other variables fixed we update S and Z as

$$S^{j+1} = \text{prox}_{\frac{\lambda}{\beta} \|\cdot\|_1} (Y - B^{j+1} - Z^j + \tilde{Z}) \quad (3)$$

$$Z^{j+1} = \text{prox}_{\frac{1}{\beta} \|\cdot\|_1} (Y - B^{j+1} - S^{j+1} + \tilde{Z}) \quad (4)$$

Where prox is the soft thresholding operator. The convergence conditions are defined as: $\|B^{j+1} - B^j\|_\infty < \varepsilon$, $\|S^{j+1} - S^j\|_\infty < \varepsilon$, $\|Z^{j+1} - Z^j\|_\infty < \varepsilon$, $\|\tilde{Z}^{j+1} - \tilde{Z}^j\|_\infty < \varepsilon$. Here ε is a predefined tolerance. Next we apply a classifier such as k-means on B^{j+1} to find the desired k clusters. The column indices of the bands closest to the cluster centroids are stored in the set Q . Then the corresponding bands from the original matrix Y represent the desired band subset. The entire algorithm is presented in Algorithm 1.

Algorithm 1: Hyperspectral Band Selection Based on Matrix CUR Decomposition

Input: $Y \in \mathbb{R}^{\ell \times n}$, maximum number of iterations T , number of sampled rows and columns s_r and s_c , number of desired bands k , parameters $\gamma_1, \gamma_2, \beta$, and tolerance ε

Output: The index Q of the desired band set.

1. Construct the symmetrically normalized spectral and spatial graph Laplacians, Φ_c and Φ_s .

2. Optimize the model in (1) using ADMM:

Initialize: $B^0, S^0, Z^0 = 0$

for $j = 0, 1, 2, 3, \dots, T - 1$ **do**

 Update B^{j+1} as in (2)

 Update S^{j+1} by solving (3)

 Update Z^{j+1} by solving (4)

 Update $\tilde{Z}^{j+1} = \tilde{Z}^j + Y - B^{j+1} - S^{j+1} - Z^{j+1}$

 Check the convergence conditions

if converged **then**

 Exit and set $\tilde{B} = B^{j+1}$

end

end

Set $\tilde{B} = B^T$

3. Cluster columns of $\tilde{B}$ using a classifier, e.g., k-means, and find the index set Q which indicates the bands closest to the k cluster centroids.

4. NUMERICAL EXPERIMENTS

4.1. Experimental Setup

In the numerical experiments, we use two benchmark HSI data sets¹: Indian Pines and Salinas-A. The Indian Pines data set was gathered by AVIRIS sensor in North-western Indiana and consists of 145×145 pixels, 200 bands, and 16 classes. The Salinas-A data set is a subscene of a larger image gathered by AVIRIS sensor in California and consists of 86×83

pixels, 204 bands, and 6 classes. See Fig. 1 for the two sample bands of the aforementioned data sets.

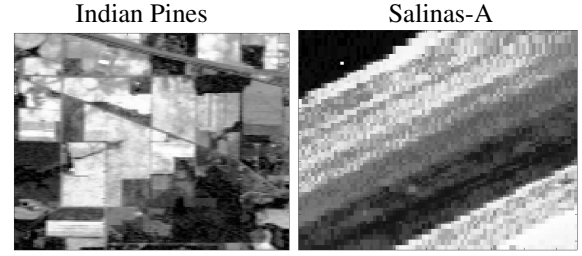


Fig. 1. Images of data sets

For comparison, we have included three state-of-the-art methods: E-FDPC [2], SR-SSIM [4], and FNGBS [3]. Here E-FDPC, SR-SSIM, and FNGBS do not require parameter tuning. FRPCALG is not being compared because it is associated with our method when $\gamma_1 = 0$, a special case within our model. However, our empirical findings have demonstrated that our method performs better when $\gamma_1 > 0$. For our method, the maximum number of iterations, the termination tolerance ε and β are set as 100, 10^{-6} , and 1 respectively. The kernel parameters σ_c, σ_s for the Laplacian graph are determined as the average distance of connected bands which differs for each data set. The number of columns s_c and rows s_r selected during the CUR decomposition step are calculated as $s_c = \text{round}(k \ln(n))$ and $s_r = \text{round}(k \ln(\ell n))$. While the number of chosen rows and columns is pre-determined, the columns and rows chosen are randomly selected using a random permutation and `rng(1)` which initializes the Mersenne Twister generator using a seed of 1. Finally, for each data set, classification method, and desired number of bands k , the parameters $\lambda, \gamma_1, \gamma_2$ and τ were tuned via a grid search. The parameter λ was tuned via a grid search in $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100, 1000\}$ while γ_1, γ_2 and τ were tuned via a grid search in $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$.

To evaluate the effectiveness of our method, classification experiments are conducted. Support vector machine (SVM) and k -nearest neighborhood (KNN) are adopted as classifiers to examine the overall accuracy (OA). For each data set, 10% of the samples are randomly selected to train the classifier. The remaining 90% are used for testing. Each experiment is repeated 50 times to reduce the randomness. The number of chosen bands varies from 3 to 30 in increments of 3. All the experiments were implemented in MATLAB 2022b on a desktop computer with Intel CPU i9-9960X RAM 64GB and GPU Dual Nvidia Quadro RTX5000 with Windows 10 Pro.

4.2. Experiment 1: Indian Pines

Fig. 2 plots the OA curves produced by SVM (left) and KNN (right) for the Indian Pines data set. The proposed method outperforms the state-of-the-art methods in terms of OA when SVM is used to classify the bands and $k \in \{3, 6, 9, 12, 15, 18, 21, 24, 27\}$. When $k = 30$ our method is comparable to FNGBS, with a marginal difference of only

¹https://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Indian_Pines

0.0065 in OA as evaluated by SVM. Similarly, when KNN is used to classify the bands, the proposed method also outperforms the state-of-the-art methods in terms of OA for $k \in \{3, 6, 9, 12, 15, 18, 21, 24, 27\}$. When $k = 30$ the proposed method performs comparably to FNGBS, where the difference in OA by KNN is only 0.0054. The average running times for each method in seconds averaged over the number of selected bands k are 0.0382 for E-FDPC, 0.1417 for FNGBS, 30.1334 for SR-SIM, and 15.3760 for the proposed method. This running time for our method is expected as it involves 100 iterations and a classification to find the band cluster centroids. The proposed method is not only faster than SR-SSIM, but also yields better classification performance.

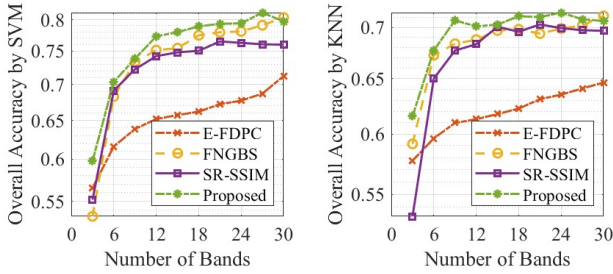


Fig. 2. Overall Accuracy for Indian Pines

4.3. Experiment 2: Salinas-A

Fig. 3 plots the OA curves produced by SVM and KNN for the Salinas data set. The proposed method performs comparably to the competitive methods. The standard deviation of the OA by SVM averaged over different k is 0.25% for E-FDPC, 0.27% for FNGBS, 0.29% for SR-SSIM, and 0.28% for the proposed method. The standard deviation of the OA by KNN averaged over different k is 0.23% for E-FDPC, FNGBS, and SR-SSIM, and 0.26% for the proposed method. Table 1 details the ranges of the standard deviation for OA in our method and thus demonstrates the stability. The average running times for each method in seconds averaged over the number of selected bands k are 0.0107 for E-FDPC, 0.0502 for FNGBS, 22.673 for SR-SSIM, and 6.1636 for the proposed method. Our method outperforms SR-SSIM in terms of computational speed.

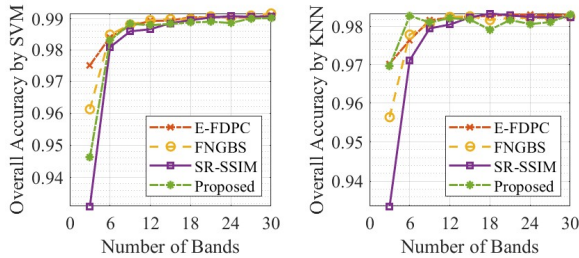


Fig. 3. Overall Accuracy for Salinas-A

Table 1. Standard Deviation Range of OA for Our Method

	Indian Pines	Salinas-A
SVM	0.0054 ~ 0.0218	0.0019 ~ 0.0029
KNN	0.0055 ~ 0.0069	0.0024 ~ 0.0033

5. CONCLUSIONS AND FUTURE WORKS

We proposed a novel method for hyperspectral band selection. Our method utilizes spatial and spectral graph regularization terms on the sparse and low-rank data with the addition of the CUR matrix decomposition to the ADMM algorithm framework. Experimental results on two real data sets demonstrate that our proposed method outperforms several state-of-the-art methods. In future work, we will explore effective strategies for applying classification methods to the lower-dimensional components generated through the CUR decomposition.

6. REFERENCES

- [1] Weiwei Sun and Qian Du, "Hyperspectral band selection: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, 2019.
- [2] Sen Jia, Guihua Tang, Jiasong Zhu, and Qingquan Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 54, no. 1, pp. 88–102, 2015.
- [3] Qi Wang, Qiang Li, and Xuelong Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5028–5039, 2021.
- [4] Buyun Xu, Xihai Li, Weijun Hou, Yiting Wang, and Yiwei Wei, "A similarity-based ranking method for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9585–9599, 2021.
- [5] Weiwei Sun and Qian Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 56, no. 6, pp. 3185–3195, 2018.
- [6] Rita Ammanouil, André Ferrari, and Cédric Richard, "A graph laplacian regularization for hyperspectral data unmixing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1637–1641.
- [7] Jing Qin, Ruilong Shen, Ruihan Zhu, and Biyun Xie, "Robust dual-graph regularized moving object detection," in *IEEE International Conference on Mechatronics and Automation (ICMA)*, 2022, pp. 487–492.
- [8] Keaton Hamm and Longxiu Huang, "Perturbations of CUR decompositions," *SIAM Journal on Matrix Analysis and Applications*, vol. 42, no. 1, pp. 351–375, 2021.