# FAST BLIND HYPERSPECTRAL UNMIXING BASED ON GRAPH LAPLACIAN

*Jing Qin*[1]*, Harlin Lee*[2]*, Jocelyn T. Chi*[3]*, Yifei Lou*[4]*, Jocelyn Chanussot*[5]*, and Andrea L. Bertozzi*[6]

[1]Department of Mathematics, University of Kentucky, Lexington, KY 40506
[2]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213
[3]Department of Statistics, North Carolina State University, Raleigh, NC 27695
[4]Department of Mathematical Sciences, University of Texas, Dallas, TX 75080
[5]University Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[6]Department of Mathematics, University of California, Los Angeles, CA 90095

## ABSTRACT

Blind hyperspectral unmixing is a challenging problem in remote sensing, which aims to infer material spectra and abundances from the given hyperspectral data. Many traditional methods suffer from poor identification of materials and/or expensive computational costs, which can be partially eased by trading the accuracy with efficiency. In this work, we propose a fast graph-based blind unmixing approach. In particular, we apply the Nyström method to efficiently approximate eigenvalues and eigenvectors of a matrix corresponding to a normalized graph Laplacian. Then the alternating direction method of multipliers (ADMM) yields a fast numerical algorithm. Experiments on a real dataset illustrate great potential of the proposed method in terms of accuracy and efficiency.

***Index Terms***— Hyperspectral imaging, hyperspectral unmixing, Nyström method, graph Laplacian, alternating direction method of multipliers.

## 1. INTRODUCTION

Hyperspectral imaging (HSI) has been widely used in remote sensing with many applications including social security, agriculture, biology, health care, and astronomy. Unlike digital images with one or three color channels, a hyperspectral image often contains hundreds or thousands of spectral bands at each recorded pixel to facilitate clustering and classification. Unfortunately, due to low spatial resolution, it is difficult in HSI to separate materials, some of which may jointly occupy at a single pixel. Hence, hyperspectral unmixing (HSU) aims at decomposing a hyperspectral image into a linear combination of spectra of pure materials (also known as *endmembers*), in which the linear coefficients correspond to the proportions or *abundances* of each pure material in a mixed pixel [1]. A more sophisticated process that estimates abundances and endmember signatures simultaneously from the HSI data is called *blind hyperspectral unmixing*.

There are a large number of HSU methods based on geometrical, statistical, and/or variational modeling of the problem. For example, it is physically reasonable to assume that all the endmembers and abundances are nonnegative, and hence nonnegative matrix factorization (NMF) [2] is one of the most popular methods due to its simple formulation and fast computation. However, the nonconvex nature of the unmixing problem leads to many local minimizers, and thereby yields poor identification of materials. Some regularization techniques including $\ell_1$-norm [3], $\ell_0$-norm [4], and total variation (TV) [5, 6] have been applied to HSU in attempts to preserve spatial smoothness of abundances or to promote joint spatial-spectral sparsity. Recently, graph-based regularization has attracted tremendous interest [7, 8, 9]. When representing hyperspectral data as a graph, each spectrum vector is considered as a node in the graph, whose affinity matrix encodes the pairwise similarities of nodes. Due to the linear relationship between spectra and abundances, abundance maps at two pixels are similar to each other if their corresponding spectra are similar. In other words, abundance maps inherit the graph structure from the spectra data. However, pairwise similarity is typically a computational bottleneck for many graph-based algorithms, especially when the HSI data is of high dimension.

In this paper, we propose an efficient way to incorporate a graph regularization for blind hyperspectral unmixing. In particular, we apply the Nyström method [10] to approximate the eigenvalues and eigenvectors of a normalized graph Laplacian, constructed from the given hyperspectral data. In addition to the nonnegative constraint for both endmembers and abundances, we assume that the sum of abundances at each pixel is one. In order to solve the constrained graph-based

The initial research for this effort was conducted at the Research Collaboration Workshop for Women in Data Science and Mathematics that was held at ICERM, July 29-August 2, 2019. Funding for the workshop was provided by ICERM. In addition, Qin is supported by the NSF grant DMS-1941197, Lou by the NSF CAREER grant DMS-1846690, Lee by the ONR grant N00014-19-1-2404, Chi by the NSF grant DMS-1760374, Chaussot by the Grant ANR-16 ASTR-0027-01, and Bertozzi by the DARPA grant FA8750-18-2-0066

unmixing model, we apply the alternating direction method of multipliers (ADMM) [11]. Motivated by an ADMM approach for solving the NMF problem [12, 13], we introduce two auxiliary variables to deal with the linear constraints, i.e., nonnegativity and sum-to-one. As a consequence, each subproblem can be solved efficiently with a closed-form solution. Numerical experiments on a real dataset show that our method yields reasonable performance with high efficiency.

The rest of the paper is organized as follows. In Section 2, we provide background knowledge including graph construction and the Nyström method. Section 3 presents details of the proposed algorithm. Experiments are conducted in Section 4, followed by conclusions and future works in Section 5.

## 2. BACKGROUND

In this section, we present how to construct a graph corresponding to the given hyperspectral data as well as how to apply the Nyström method to approximate the eigenvalues and eigenvectors of the graph Laplacian.

Given a collection of spectral vectors $\mathcal{V} = \{\mathbf{x}_i\}_{i=1}^n \subseteq \mathbb{R}^w$ with $n$ being the number of pixels in the hyperspectral data, we define an affinity matrix, or similarity matrix, $W \in \mathbb{R}^{n\times n}$ of the underlying graph as

$$W_{ij} = e^{-d(\mathbf{x}_i,\mathbf{x}_j)^2/\sigma}, \quad i,j = 1,\dots,n, \tag{1}$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between the two spectral vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\sigma > 0$ controls how similar they are. Following [14], we adopt the cosine similarity as the distance function for HSI, i.e.,

$$d(\mathbf{x}_i, \mathbf{x}_j) = 1 - \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}.$$

Calculating pairwise similarities of a fully-connected graph is the crux of many graph-based algorithms. In order to reduce the computational cost, we apply the Nyström method [10] to approximate the eigenvectors and eigenvalues of $W$ by using a small number of sampled data points. Up to permutations, the similarity matrix $W$ can be expressed in a block-matrix form,

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix},$$

where $W_{11}$ is the similarity matrix of the sampled points, $W_{12} = W_{21}^T$ is the one of the sampled points and the unsampled points, and $W_{22}$ is the one of the unsampled points. Assume that the symmetric matrix $W_{11}$ has the eigendecomposition: $W_{11} = U\Lambda U^T$, where $U$ has orthonormal eigenvectors as columns and $\Lambda$ is a diagonal matrix whose diagonal entries are eigenvalues of $W_{11}$. The Nyström extension gives an approximation of $W$ by using $U$ and $\Lambda$ as follows,

$$W \approx \widehat{U}\Lambda\widehat{U}^T, \quad \text{where} \quad \widehat{U} = \begin{bmatrix} U \\ W_{21}U\Lambda^{-1} \end{bmatrix}. \tag{2}$$

In other words, computation of the pairwise similarity matrix $W$ can be significantly reduced by using a small set of sampled points.

It has been shown in [15, 16] that a normalized similarity matrix yields better performance with more efficient computation. Therefore, we consider to normalize the weight $W$ as,

$$\widetilde{W} = D^{-1/2} W D^{-1/2}, \tag{3}$$

where $D$ is called the degree matrix, i.e., a diagonal matrix with column sums of $W$ as its diagonal entries. Similarly, $\widetilde{W}$ can be approximated via (2), i.e., $\widetilde{W} \approx V\widetilde{\Lambda}V^T$, where $V \in \mathbb{R}^{n\times d}$ and the diagonal elements of $\Lambda$ are eigenvectors and eigenvalues of the approximated weight by using $d$ ($d \ll n$) sampled points. Denoting the graph Laplacian by $L := I - \widetilde{W}$ with the identity matrix $I$, we have the eigendecomposition form of $L = V\Sigma V^T$, where $\Sigma = I - \widetilde{\Lambda}$ and $V$ is the same as that in the eigendecomposition of $\widetilde{W}$. Please refer to [17] for more details.

## 3. PROPOSED METHOD

Consider a hyperspectral data $X = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{w\times n}$, where $w$ is the number of wavelengths and $n$ is the number of spatial pixels. We assume that the spectral measurement at each pixel is a linear combination of the endmember's spectra. Suppose there are $k$ pure materials to be considered and we denote a matrix $S \in \mathbb{R}^{w\times k}$ as a dictionary of endmembers' spectra, each column representing one material. The linear coefficients or the abundances that represent the pixel $\mathbf{x}_i$ over this dictionary $S$ is denoted as $\mathbf{a}_i$. By organizing $\mathbf{a}_i$ as a column vector, we obtain a matrix $A \in \mathbb{R}^{k\times n}$. In short, we assume the hyperspectral data can be modelled as

$$X = SA + \varepsilon,$$

where $\varepsilon \in \mathbb{R}^{w\times n}$ is a noise term, which is often assumed to have a Gaussian distribution.

The blind hyperspectral unmixing is to recover two matrices $S$ and $A$, given the hyperspectral data $X$. As it is highly ill-posed, additional assumptions and proper regularizations are necessary. In addition to the standard constraints of nonnegativity and sum-to-one, we consider a graph-based regularization on the abundance matrix $A$ formulated as follows,

$$J(A) = \frac{1}{2}\sum_{i,j=1}^{n} \|\mathbf{a}_i - \mathbf{a}_j\|^2 \widetilde{W}_{ij}, \tag{4}$$

where $\widetilde{W}_{ij}$ is a normalized weight between pixels $\mathbf{x}_i$ and $\mathbf{x}_j$, defined in (3). By minimizing the regularization term $J(A)$, we assume that two column vectors of $\mathbf{a}_i, \mathbf{a}_j$ in abundance matrix should be close to each other if the hyperspectral measurements $\mathbf{x}_i, \mathbf{x}_j$ are similar. Simple calculations show that

$$J(A) = \sum_{i=1}^{n} \mathbf{a}_i^T\mathbf{a}_i - \sum_{i,j=1}^{n} \mathbf{a}_i^T\mathbf{a}_j\widetilde{W}_{ij} = \mathrm{tr}(ALA^T),$$

where $\mathrm{tr}(\cdot)$ is the trace operator summing up all diagonal entries of a matrix.

Now we are ready to formulate the proposed blind unmixing model,

$$\min_{\substack{S,A\geq \mathbf{0}\\ \mathbf{1}_k^T A=\mathbf{1}_n^T}} \frac{1}{2}\|X-SA\|_F^2+\frac{\lambda}{2}\,\mathrm{tr}(ALA^T), \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\mathbf{1}_k$ stands for the column vector whose entries are all ones. The constraint $\mathbf{1}_k^T A=\mathbf{1}_n^T$ means that all columns of $A$ belong to the probability simplex, i.e., the set of any nonnegative vector that sums to one.

We introduce two indicator functions to deal with the constraint sets in (5). Generally, we define the indicator function $\mathbb{1}_G$ of a set $G\subseteq\mathbb{R}^n$ as

$$\mathbb{1}_G(Z)=\begin{cases}0, & Z\in G;\\ \infty, & \text{otherwise.}\end{cases}$$

Denote $P:=\mathbb{R}_+^{w\times k}$ be the set of all nonnegative matrices of the size $w\times k$ and the set $N:=\{Z\in\mathbb{R}^{k\times n}: Z\geq \mathbf{0}, \mathbf{1}_k^T Z=\mathbf{1}_n^T\}$. Then we can rewrite the model (5) as,

$$\min_{S,A}\frac{1}{2}\|X-SA\|_F^2+\frac{\lambda}{2}\,\mathrm{tr}(ALA^T)+\mathbb{1}_P(S)+\mathbb{1}_N(A). \tag{6}$$

In order to apply the ADMM framework to minimize (6), we further introduce two auxiliary variables $B,C\in\mathbb{R}^{k\times n}$. Specifically, we split variables and rewrite (6) into an equivalent form,

$$\begin{aligned}\min_{S,A,B,C}\ &\frac{1}{2}\|X-SA\|_F^2+\frac{\lambda}{2}\,\mathrm{tr}(BLB^T)+\mathbb{1}_P(C)+\mathbb{1}_N(A)\\ \text{s.t.}\quad & A=B,\ S=C.\end{aligned}$$

The augmented Lagrange function is then given by

$$\begin{aligned}\mathcal{L}=&\frac{1}{2}\|X-SA\|_F^2+\frac{\lambda}{2}\,\mathrm{tr}(BLB^T)+\mathbb{1}_P(C)\\ &+\mathbb{1}_N(A)+\frac{\rho}{2}\|A-B+\widetilde{B}\|_F^2+\frac{\gamma}{2}\|S-C+\widetilde{C}\|_F^2,\end{aligned}$$

where $\widetilde{B},\widetilde{C}$ are dual variables and $\rho,\gamma$ are two positive parameters. Then ADMM yields the following algorithm

$$\begin{cases}S\leftarrow \underset{S}{\mathrm{argmin}}\ \frac{1}{2}\|X-SA\|_F^2+\frac{\gamma}{2}\|S-C+\widetilde{C}\|_F^2\\ A\leftarrow \underset{A\in\Pi}{\mathrm{argmin}}\ \frac{1}{2}\|X-SA\|_F^2+\frac{\rho}{2}\|A-B+\widetilde{B}\|_F^2\\ B\leftarrow \underset{B}{\mathrm{argmin}}\ \frac{\lambda}{2}\,\mathrm{tr}(BLB^T)+\frac{\rho}{2}\|A-B+\widetilde{B}\|_F^2\\ C\leftarrow \underset{C\geq\mathbf{0}}{\mathrm{argmin}}\ \frac{\gamma}{2}\|S-C+\widetilde{C}\|_F^2\\ \widetilde{B}\leftarrow \widetilde{B}+A-B\\ \widetilde{C}\leftarrow \widetilde{C}+S-C.\end{cases} \tag{7}$$

For the $S$-subproblem in (7), the Karush-Kuhn-Tucker (KKT) condition indicates that $-XA^T+SAA^T+\gamma(S-C+\widetilde{C})=\mathbf{0}$, leading to a closed-form solution for $S$, i.e.,

$$S=\big(XA^T+\gamma(C-\widetilde{C})\big)(AA^T+\gamma I)^{-1}. \tag{8}$$

Note $AA^T\in\mathbb{R}^{k\times k}$ has a small matrix size as $k\ll n$, which implies that the $S$-subproblem is fast to solve.

As for the $A$-subproblem, we adopt the fast algorithm in [18] that involves the projection onto the set $N$, denoted by $\Pi_N$. The KKT condition gives a closed-form solution for $A$,

$$A=\Pi_N\Big((S^TS+\rho I)^{-1}\big(S^TX+\rho(B-\widehat{B})\big)\Big). \tag{9}$$

The $B$-subproblem involves the graph Laplacian. As detailed in Section 2, we can approximate $L$ by $V\Sigma V^T$. The KKT condition for the $B$-subproblem is (7) is given by

$$BV\Sigma V^T+\mu(B-A-\widetilde{B})=\mathbf{0}, \tag{10}$$

where $\mu=\rho/\lambda$ and hence we can solve for $B$ by

$$B=\mu(A+\widetilde{B})V(\Sigma+\mu I)^{-1}V^T. \tag{11}$$

Notice that the matrix to be inverted is a diagonal matrix of size $d\times d$, which enjoys fast computation.

Finally, we have the closed-form solution for the $C$-subproblem,

$$C=\max(S+\widetilde{C},\mathbf{0}), \tag{12}$$

which is an element-wise operation to project onto the nonnegative set. The entire algorithm is summarized in Algorithm 1. The stopping criteria are to set $\|S_i-S_{i+1}\|_F/\|S_i\|_F$ and $\|A_i-A_{i+1}\|_F/\|A_i\|_F$ smaller than some tolerance.

**Algorithm 1** Blind Hyperspectral Image Unmixing Based on the Graph Laplacian

**Input:** The data $X$, parameters $\rho$, $\lambda$, and maximum number of iterations $T$, tolerance $tol$.
**Output:** $S$ and $A$.
**Initialize:** $S_0$, $A_0$ and use Nyström method to get the reduced eigendecomposition form of the graph Laplacian $L=V\Sigma V^T$.
**for** $t=0,\ldots,T-1$ **do**
  Update $S_{t+1}$ via (8).
  Update $A_{t+1}$ via (9).
  Update $B_{t+1}$ via (11).
  Update $C_{t+1}$ via (12).
  Set $\widetilde{B}_{t+1}=\widetilde{B}_t+(A_{t+1}-B_{t+1})$.
  Set $\widetilde{C}_{t+1}=\widetilde{C}_t+(S_{t+1}-C_{t+1})$.
  Stop if the stopping criteria are met.
**end for**

## 4. NUMERICAL RESULTS

We conduct numerical experiments on a real hyperspectral dataset, called Urban[1], which has $307 \times 307$ pixels and 162 spectral bands. The ground truth that consists of six identified endmember labels and their corresponding abundances is shown on the top row of Fig. 1. We compare the proposed method, denoted by GraphL, with two competing methods: fully constrained least squared unmixing (FCLSU) [19] and a recent work of fractional norm penalty method with $q = 0.1$, denoted by FRAC [20] . All experiments are performed in MATLAB 2018b on a MacBook Pro 2017 with an 2.9 GHz Intel Core i7 and 16GB RAM in double precision.

To quantitatively measure the performance, we adopt the following two metrics to calculate the error between an estimator $\widehat{Y} \in \mathbb{R}^{r\times c}$ and the ground truth $Y \in \mathbb{R}^{r\times c}$:

(a) Root-mean-square error (RMSE)

$$RMSE(Y,\widehat{Y}) = \frac{1}{c}\sqrt{\frac{1}{r}\sum_{i=1}^{r}\|\mathbf{y}_i - \widehat{\mathbf{y}}_i\|_2^2},$$

where $\mathbf{y}_i \in \mathbb{R}^c$ is the $i$-th row of $Y$;

(b) Normalized mean-square error (nMSE)

$$nMSE(Y,\widehat{Y}) = \frac{\|Y-\widehat{Y}\|_F}{\|Y\|_F}.$$

In order to make a fair comparison, we use the initialization steps in [20]. In particular, we run vertex component analysis (VCA) [21], which results in 60 endmember candidates that are clustered into 6 groups. This is directly used as $S$ for FCLSU and FRAC, while we use the mean spectrum within each group and the sum of the abundances estimated by FCLSU within each group as initial conditions for $S_0$ and $A_0$, respectively. We randomly sample $0.5\%$ of the pixels for the Nyström method to approximate the graph Laplacian, and use $\sigma = 5$ in (1). When choosing $\rho$ and $\lambda$, we perform a grid search with parameter candidates evenly spaced over the interval in a logarithmic spacing, i.e., $\rho \in \{10^{-3}, 10^{-2.6}\ldots, 10^{0.6}, 10^{1}\}$, $\mu = \rho/\lambda \in \{10^{-6},\ldots 1\}$, and $\gamma \in \{10^{-2}\ldots, 10, 10^2, 10^3\}$ for GraphL. For the FRAC method, we fix $\rho = 10$ as suggested in [20] and choose $\lambda$ among $\{10^{-3}, 10^{-2.6}\ldots, 10^{0.6}, 10^{1}\}$. The optimal parameters are chosen based on visual inspection of the resulting abundance vectors, and are summarized in Table 1.

The quantitative comparisons in Table 1 indicate that GraphL achieves the best results in terms of RMSE and nMSE. It also requires less computational time than FRAC. Fig. 1 shows that the GraphL produces the abundances that are visually similar to those from FCLSU and FRAC. But the third panel (Tree) of GraphL looks much closer to the ground truth. Overall, GraphL has great potential in hyperspectral unmixing, especially for high-dimensional data.

[1]The data is downloaded from `http://www.escience.cn/people/feiyunZHU/Dataset_GT.html`

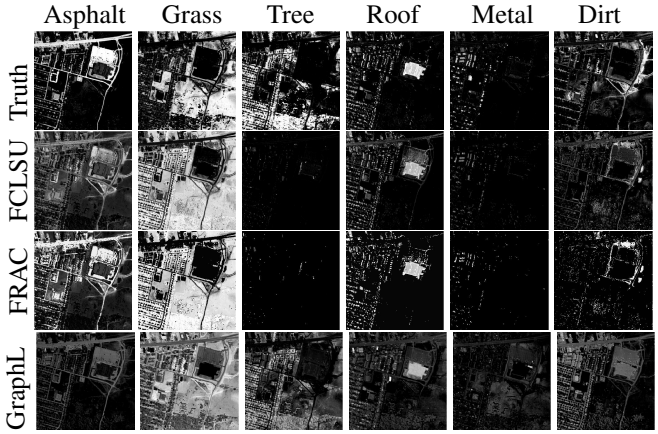


**Fig. 1**. Abundance matrices ($A$) of the Urban data produced by FCLSU, FRAC, and GraphL. Top row is the ground truth.

| | **FCLSU** | **FRAC** | **GraphL** |
|---|---|---|---|
| $RMSE(S,\hat{S})$ | 0.196 | * | **0.151** |
| $nMSE(S,\hat{S})$ | 0.807 | * | **0.675** |
| $RMSE(A,\hat{A})$ | 0.240 | 0.251 | **0.207** |
| $nMSE(A,\hat{A})$ | 0.849 | 0.930 | **0.681** |
| time (sec) | **34** | 119 | $74^{\dagger}+5$ |
| $\lambda$ | n/a | $10^{0.2}$ | $10^{3}$ |
| $\rho$ | n/a | 10 | $10^{-3}$ |
| $\gamma$ | n/a | n/a | $10^{3}$ |
| iterations | n/a | 200 | 50 |

$*$ : same as FCLSU, since FRAC only estimates $A$.
$\dagger$ : time spent estimating the graph Laplacian matrix $L$.

**Table 1**. Quantitative comparison of the unmixing performances and a summary of the chosen parameters. The best results in each row are highlighted in bold.

## 5. CONCLUSIONS

In this paper, we proposed a blind hyperspectral unmixing model based on a normalized graph Laplacian. To enhance the computational efficiency for high-dimensional hyperspectral data, we adopted the Nyström method to approximate the eigenvalues and eigenvectors of the graph Laplacian. By introducing auxiliary variables, we applied ADMM to minimize the proposed model in a way that each subproblem has a closed-form solution. Experiments on a real dataset have shown promising results of the proposed method in terms of efficiency and identification accuracy. Future works include convergence analysis and comprehensive experiments on both synthetic and real datasets in comparison with the state-of-

the-art methods in blind hyperspectral unmixing.

## 6. REFERENCES


[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[3] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, 2016.

[4] M-D Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.

[5] Z. Guo, T. Wittman, and S. J. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*. International Society for Optics and Photonics, 2009, vol. 7334, p. 73341M.

[6] M-D Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, 2012.

[7] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2010.

[8] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, 2012.

[9] R. Ammanouil, A. Ferrari, and C. Richard, "A graph Laplacian regularization for hyperspectral data unmixing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 1637–1641.

[10] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nyström method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 214–225, 2004.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[12] Y. Zhang, "An alternating direction algorithm for nonnegative matrix factorization," Tech. Rep., 2010.

[13] J. Qin, T. Laurent, K. Bui, R. V. Tan, J. Dahilig, S. Wang, J. Rohe, J. Sunu, and A. L. Bertozzi, "Detecting plumes in LWIR using robust nonnegative matrix factorization with graph-based initialization," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXI*. International Society for Optics and Photonics, 2015, vol. 9472, p. 94720V.

[14] H. Hu, J. Sunu, and A. L. Bertozzi, "Multi-class graph Mumford-Shah model for plume detection using the MBO scheme," in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2015, pp. 209–222.

[15] A. L. Bertozzi and A. Flenner, "Diffuse interface models on graphs for classification of high dimensional data," *Multiscale Model. Simul*, vol. 10, no. 3, pp. 1090–1118, 2012.

[16] E. Merkurjev, J. Sunu, and A. L Bertozzi, "Graph mbo method for multiclass segmentation of hyperspectral stand-off detection video," in *IEEE International Conference on Image Processing*, 2014, pp. 689–693.

[17] Z. Meng, E. Merkurjev, A. Koniges, and A. L Bertozzi, "Hyperspectral image classification using graph clustering methods," *Image Processing On Line*, vol. 7, pp. 218–245, 2017.

[18] W. Wang and M. A. Carreira-Perpinán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *arXiv preprint arXiv:1309.1541*, 2013.

[19] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.

[20] L. Drumetz, T. R. Meyer, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3435–3450, 2019.

[21] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.