Open access • Journal Article • DOI:10.1109/TGRS.2020.3020810

Blind Hyperspectral Unmixing Based on Graph Total Variation Regularization

Jing Qin, Harlin Lee, Jocelyn T. Chi, Lucas Drumetz ...+3 more authors

Institutions: University of Kentucky, Carnegie Mellon University, North Carolina State University, Grenoble Institute of Technology ...+2 more institutions

Published on: 01 Apr 2021 - IEEE Transactions on Geoscience and Remote Sensing (IEEE)

Topics: Hyperspectral imaging, Total variation denoising, Graph (abstract data type) and Pixel

Blind Hyperspectral Unmixing Based on Graph Total Variation Regularization

Jing Qin, *Member, IEEE,* Harlin Lee, *Student Member, IEEE,* Jocelyn T. Chi, Lucas Drumetz, *Member, IEEE,* Jocelyn Chanussot, *Fellow, IEEE,* Yifei Lou, *Member, IEEE,* and Andrea L. Bertozzi, *Member, IEEE*

***Abstract*—Remote sensing data from hyperspectral cameras suffer from limited spatial resolution, in which a single pixel of a hyperspectral image may contain information from several materials in the field of view. Blind hyperspectral image unmixing is the process of identifying the pure spectra of individual materials (i.e., endmembers) and their proportions (i.e., abundances) at each pixel. In this paper, we propose a novel blind hyperspectral unmixing model based on the graph total variation (gTV) regularization, which can be solved efficiently by the alternating direction method of multipliers (ADMM). To further alleviate the computational cost, we apply the Nyström method to approximate a fully-connected graph by a small subset of sampled points. Furthermore, we adopt the Merriman-Bence-Osher (MBO) scheme to solve the gTV-involved subproblem in ADMM by decomposing a grayscale image into a bit-wise form. A variety of numerical experiments on synthetic and real hyperspectral images are conducted, showcasing the potential of the proposed method in terms of identification accuracy and computational efficiency.**

***Index Terms*—Blind hyperspectral unmixing, Nyström method, graph Laplacian, graph total variation, alternating direction method of multipliers.**

I. Introduction

Hyperspectral imaging (HSI) is an important and useful tool to acquire high resolution data in the electromagnetic spectrum with many applications in remote sensing, including surveillance, agriculture, environmental monitoring, and astronomy. With hundreds to thousands of spectral bands, a hyperspectral image provides a detailed description of a scene. However, due to limited spatial resolution of imaging sensors, the acquired hyperspectral data at each pixel represents a collection of material signatures in the field of view of each pixel. The signature corresponding to one pure material is called an *endmember* in hyperspectral data analysis [1]. Given the endmembers of all materials present in the scene, hyperspectral unmixing aims to estimate the proportions of constituent endmembers at each single pixel, called the *abundance map*. If the spectral information of endmembers is unavailable, then the problem becomes a blind hyperspectral unmixing problem that requires simultaneously identifying the endmembers and estimating the abundance map. There are a large number of hyperspectral mixing and unmixing methods [2], [3], including linear and nonlinear models, depending on assumptions about the interaction of the light with the observed scene.

In this paper, we focus on the linear mixing model. Specifically, by assuming that each light ray interacts with only one endmember in the field of view before reaching the sensor, we model the spectrum at each pixel as a linear combination of all endmembers. Due to the physical interpretation of the hyperspectral mixing model, it is also reasonable to assume that each element of endmembers and abundances is nonnegative. Another commonly used constraint is that abundances from all the endmembers at each pixel sum up to one, which implies that all abundance vectors belong to the probability simplex, determined by the standard unit vectors in a Euclidean space. Note that one can remove the sum-to-one constraint for physically motivated reasons, e.g., when illumination conditions or the topography of the scene change locally in the image [4]. We adopt the sum-to-one constraint due to the interpretability of the abundances.

Nonnegative matrix factorization (NMF) [5], decomposing a given matrix into a product of two matrices with nonnegative entries, is widely used in blind hyperspectral unmixing [6], [7], [8]. Suppose the given hyperspectral image X is of size $w \times n$, where w is the number of spectral bands and n is the number of spatial pixels. One aims to write X as a product of two nonnegative matrices $S \in \mathbb{R}^{w\times k}$ and $A \in \mathbb{R}^{k\times n}$ with k being the total number of the endmembers. Note that the rank of the matrix SA is at most k, and k is usually much smaller than w and n. Then the hyperspectral unmixing problem can be formulated as a nonnegative least squares problem,

$$\min_{\substack{S\in\Omega_{w\times k}\\ A\in\Omega_{k\times n}}} \frac{1}{2}\|X - SA\|_{\mathrm{F}}^2, \tag{1}$$

where $\Omega_{l\times m}$ denotes the set of all nonnegative real matrices

J. Qin is with the Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA (e-mail: jing.qin@uky.edu). Qin is supported by the NSF grant DMS #1941197. H. Lee is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail:harlin@cmu.edu). Lee is supported by the ONR grant N00014-19-1-2404 and ARO grant W911NF-18-1-0303. J. T. Chi is with the Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA (e-mail: jtchi@ncsu.edu). Chi is supported by the NSF grant DMS-1760374.L. Drumetz is with IMT Altantique, LAb-STICC, UBL, 29238 Brest, France (e-mail: lucas.drumetz@imt-atlantique.fr). Drumetz is supported by Programme National de Teledétection Spatiale (PNTS), Grant PNTS-2019-4.J. Chanussot is with the University Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: jocelyn@hi.is). Chanussot is supported by the Grant ANR-16 ASTR-0027-01.Y. Lou is with the Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080, USA (e-mail: yifei.lou@utdallas.edu). Lou is supported by the NSF CAREER grant DMS-1846690. A. L. Bertozzi is with the Department of Mathematics, University of California, Los Angeles, CA 90095, USA (e-mail: bertozzi@math.ucla.edu). Bertozzi is supported by the DARPA grant FA8750-18-2-0066 and NSF DMS-1737770.

of size $l \times m$, i.e.,

$$\Omega_{l\times m} := \{X \in \mathbb{R}^{l\times m} \mid X_{ij} \geq 0,\, i = 1, \ldots, l,\; j = 1, \ldots, m\}. \tag{2}$$

However, non-convexity of the objective function in (1) may lead to multiple local minima for NMF. To address this issue, various regularization techniques have been developed to enforce some desirable properties on the endmembers or abundance matrices. For example, methods based on the spatial sparsity of abundances include the use of the ℓ_0-norm [9], the ℓ_1-norm [10], the ℓ_2-norm in fully constrained least squares unmixing (FCLSU) [11], the $\ell_{1/2}$-norm [12], and the mixed $\ell_{p,q}$-norm for group sparsity [13].

Due to the success of the total variation (TV) [14] in the image processing community, the TV regularization has been applied to hyperspectral unmixing to preserve the piecewise constant structure of the abundance map for each material. For example, sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV) [15] involves a two-dimensional TV regularization. Other TV-based variants include TV with ℓ_1 [16], TV with sparse NMF [17], TV with nonnegative tensor factorization [18], and an improved collaborative NMF with TV (ICoNMF-TV) [19] that combines robust collaborative NMF (R-CoNMF) [20] and TV. Recently, TV is considered as a quadratic regularization promoting minimum volume in the NMF framework, referred to as NMF-QMV [21]. An extension of TV to nonlocal spatial operators [22], [23] has led to nonlocal TV being considered for the blind hyperspectral unmixing problem [24], [25]. TV has also been extended from vectors in Euclidean space to signals defined on a graph. For example, the graph TV (gTV) [26] is a special case of the p-Dirichlet form [27], [28] in graph signal processing. Some graph regularization techniques for hyperspectral imaging include graph NMF (GNMF) [29], structured sparse regularized NMF (SS-NMF) [30], graph-regularized $\ell_{1/2}$-NMF (GLNMF) [31], and graph-regularized multilinear mixing model (G-MLM) based on superpixels [32]. However, most of these graph-based approaches suffer from intensive computation, especially when computing the pairwise similarity between all pixels. To reduce the computational cost, the Nyström method [33] generates a low-rank approximation of the graph Laplacian, which can be incorporated into unmixing.

In this work, we propose an efficient framework for blind hyperspectral unmixing based on an approximation of gTV to exploit the similarity of spectral information at different pixels and preserve sharp edges of the abundance map. By treating the spectral vector at each pixel as a vertex, the given hyperspectral data can be modeled as a graph, whose adjacency matrix is determined by the pairwise similarity between any two vertices. Instead of using the incidence matrix to define the discrete graph derivative operator and thereby graph TV [27], [28], [34], [26], we approximate gTV by the graph Laplacian. This approach is inspired by a theoretical result in [35]: the TV semi-norm of a binary function defined on a graph is well-approximated by the graph Ginzburg-Landau (GL) functional involving the graph Laplacian and a double-well potential. In order to relax the restriction on binary data, we adopt a bitwise decomposition [36] to deal with grayscale images. Specifically, we decompose the input data into eight bits, solve the optimization problem at each bit channel, and aggregate all bits into grayscale values.

Our framework incorporates several techniques to increase the computational efficiency. To avoid a direct calculation of the graph Laplacian, we adopt the Nyström method [33] in graph clustering to approximate the eigenvalues and eigenvectors of the graph Laplacian. The Nyström method is a low-rank approximation of the weight-matrix that does not require the computation of all pairwise comparisons between feature vectors. Rather, it uses random sampling to construct a low rank approximation that is roughly $O(N)$ for the number of feature vectors rather than computing the full matrix which is $O(N^2)$. This is a reasonable assumption in cases where the image is thought to be representable by a relatively small number of features as would be the case with a modest number of endmembers. This approximation significantly reduces the computational costs in both time and storage, which makes our approach scalable to high-dimensional data. Moreover, we design an efficient numerical algorithm to solve the proposed model via the alternating direction method of multipliers (ADMM) [37], [38]. In particular, the gTV-related subproblem can be solved efficiently by the Merriman-Bence-Osher (MBO) scheme [39], [40] at each bit channel. We can readily incorporate an accelerated version [41] of the MBO scheme and the Nyström method into the proposed framework. To demonstrate the effectiveness of these approximations, we conduct extensive experiments on various synthetic and real hyperspectral datasets, showing the great potential of the proposed method in terms of accuracy and computational efficiency.

The main contributions of this paper are three-fold:

1) We propose a novel data-driven type of graph regularization, i.e., graph TV based on the similarity of spectral information, imposed on the abundance map. To the best of our knowledge, this is the first time that the graph total variation regularization has been applied to solve a hyperspectral unmixing problem.
2) We apply the Nyström method to efficiently approximate eigenvalues and eigenvectors of a normalized graph Laplacian, which significantly improves the scalability of our approach.
3) We present an effective graph-based framework that integrates the Nyström method and the MBO scheme into blind hyperspectral unmixing. We also provide a thorough discussion of computational complexity and parameter selection of the proposed algorithm.

The remainder of the paper is organized as follows. In Section II, we provide a brief introduction of concepts and methods used in our workflow, including the Nyström method, the GL functional, and the MBO scheme. Section III presents the proposed hyperspectral unmixing model, followed by a detailed description of the proposed algorithm based on ADMM and its complexity analysis. Extensive experiments are provided in Section IV, followed by a discussion on parameter selection in Section V. Finally, conclusions and future works

are given in Section VI.

II. Preliminaries

In this section, we provide preliminary knowledge for a set of building blocks that are used in this work, including the graph construction, the Nyström method for efficiently approximating the similarity weight matrix, and the GL functional with a fast solver to find its minimizer via MBO.

A. Graph Construction

Similarity graphs are an important mathematical tool to describe directed/undirected pairwise connections between objects. Typically, a graph consists of *vertices* (or *nodes*) connected by *edges* with the associated *weights*. Consider a collection of data points $\{\mathbf{x}_i\}_{i=1}^n \subseteq \mathbb{R}^w$, one simple way to construct a graph G is to treat each point as a vertex of the graph. Then the *weight matrix* (also known as the *affinity matrix*) $W \in \mathbb{R}^{n\times n}$ of G is defined by

$$W_{ij} = e^{-d(\mathbf{x}_i,\mathbf{x}_j)^2/\sigma}, \quad i,j = 1,\dots,n, \tag{3}$$

where $d(\mathbf{x}_i,\mathbf{x}_j)$ is the distance between the two vertices $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\sigma > 0$ controls how similar they are. There are two distance metrics widely used in graph-based applications:

1) Euclidean distance: $d(\mathbf{x}_i,\mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$;
2) cosine similarity: $d(\mathbf{x}_i,\mathbf{x}_j) = 1 - \frac{\langle \mathbf{x}_i,\mathbf{x}_j\rangle}{\|\mathbf{x}_i\|_2\|\mathbf{x}_j\|_2}$.

In this paper, we adopt the cosine similarity as the distance function for hyperspectral data, which is physically motivated by the fact that illumination effects change the scaling of spectra but not their overall shape in the spectral domain [40], [42], [41].

Based on the weight matrix W, we define the degree matrix, denoted by D, as a diagonal matrix whose entries are the row (or column) sums of W. There are several ways to define graph Laplacian. For example, the standard graph Laplacian is defined as $L = D - W$, while the *(symmetric) normalized graph Laplacian* is given by

$$L_s = I - D^{-1/2}WD^{-1/2}. \tag{4}$$

In this work, we adopt the symmetric normalized graph Laplacian due to its outstanding performance in the graph-based data classification [40], [43]. By denoting $X = [\mathbf{x}_1,\dots,\mathbf{x}_n] \in \mathbb{R}^{w\times n}$, we have

$$\langle X^\top, L_s X^\top\rangle = \sum_{i,j=1}^{n} \|\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j\|_2^2 W_{ij}, \tag{5}$$

where $\widehat{\mathbf{x}}_i = \mathbf{x}_i/\sqrt{d_{ii}}$ with d_{ij} being the (i,j)-th entry of the matrix D. Here, we use the standard inner product on matrices, i.e., $\langle X^\top, L_s X^\top\rangle = \mathrm{tr}(XL_sX^\top)$, where $\mathrm{tr}(\cdot)$ is the matrix trace operator that returns the sum of all the diagonal elements.

B. Nyström Method

Computing and storing pairwise similarities of a fully-connected graph is usually a bottleneck of many graph-based algorithms. In order to reduce the time/space complexity, we apply the Nyström method [33] to approximate the eigenvalues and eigenvectors of $W \in \mathbb{R}^{n\times n}$ by using only p sampled data points with $p \ll n$. 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 $W_{11} \in \mathbb{R}^{p\times p}$ is the similarity matrix of the sampled data points, $W_{12} = W_{21}^\top$ 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 $W_{11} = U\widetilde{\Lambda}U^\top$, where U has orthonormal eigenvectors as columns and $\widetilde{\Lambda}$ is a diagonal matrix whose diagonal entries are eigenvalues of W_{11}. The Nyström extension gives an approximation of W by using U and $\widetilde{\Lambda}$ as follows,

$$W \approx \widetilde{U}\widetilde{\Lambda}\widetilde{U}^\top, \quad \text{where} \quad \widetilde{U} = \begin{bmatrix} U \\ W_{21}U\widetilde{\Lambda}^{-1} \end{bmatrix}. \tag{6}$$

Note that the columns of $\widetilde{U}$ require further orthogonalization. See [33], [41] for more details.

In this work, we apply the Nyström method to calculate the weight matrix for the sampled data and then use the approximated eigendecomposition (6) to approximate the normalized graph Laplacian, i.e.,

$$L_s \approx D^{-1/2}\widetilde{U}(I - \widetilde{\Lambda})\widetilde{U}^\top D^{-1/2} := V\Lambda V^\top, \tag{7}$$

where $V = D^{-1/2}\widetilde{U} \in \mathbb{R}^{n\times p}$ and $\Lambda = I - \widetilde{\Lambda} \in \mathbb{R}^{p\times p}$. In this way, computation of pairwise similarities is significantly reduced from the whole dataset to a small portion.

C. Ginzburg-Langdau Functional and MBO Scheme

The classic Ginzburg-Landau (GL) energy [43], [44] for diffuse interface models is

$$\frac{\epsilon}{2}\int_\Omega |\nabla u|^2 dx + \frac{1}{\epsilon}\int_\Omega \Phi(u)dx,$$

where $\Phi(u) := \frac{1}{4}u^2(u-1)^2$ is a double-well potential to enforce u to take binary values of $\{0,1\}$ on a domain Ω. The term "diffuse interface" refers to a smooth transition between two phases of u, where the smoothness is modeled by the H_1-semi norm and the scale of the transition is controlled by the parameter $\epsilon > 0$. It is proven in [45] that the GL functional Γ-converges to the TV semi-norm, i.e., as $\epsilon \to 0$,

$$\frac{\epsilon}{2}\int_\Omega |\nabla u|^2 dx + \frac{1}{\epsilon}\int_\Omega \Phi(u)dx \to C\int_\Omega \|\nabla u\| dx,$$

for some constant $C > 0$.

In a series of works including [40], [42], [46], [47], [48], the GL functional has been extended to graphs, defined as

$$GL(\mathbf{u}) = \epsilon\langle \mathbf{u}, L\mathbf{u}\rangle + \frac{1}{\epsilon}\Phi(\mathbf{u}), \tag{8}$$

where $\mathbf{u} = [u_1,\dots,u_n]^\top \in \mathbb{R}^n$ is a signal defined on a graph $\mathcal{G}$ with u_i being the state of vertex i and L is the graph Laplacian of $\mathcal{G}$ or its variant. Here $\Phi(\mathbf{u}) = \sum_{i=1}^n \Phi(u_i)$, which can be extended to the matrix case, i.e., $\Phi(U) = \sum_{i,j}\Phi(u_{ij})$ for any matrix $U = (u_{ij})$. Thanks to the double-well potential, the GL functional has been successfully applied to binary data classification [40] and multiclass classification [41], [46]. We

employ the binary model here. By adding a fidelity term to the GL energy, one obtains the following minimization problem

$$E(\mathbf{u}) = GL(\mathbf{u}) + \lambda F(\mathbf{u}), \tag{9}$$

where $F(\mathbf{u})$ is a differentiable functional that fits the unknown variable $\mathbf{u}$ to the given data $\mathbf{y}$, e.g., $F(\mathbf{u},\mathbf{y}) = \frac{1}{2}\|\mathbf{u}-\mathbf{y}\|_2^2$. The parameter $\lambda > 0$ balances the contributions between the GL regularization term and the data fidelity term. When $\mathbf{u}$ is binary, the energy E can be efficiently minimized via the MBO scheme [39], [40]. In particular, the MBO scheme alternates a gradient descent step that minimizes $\langle \mathbf{u}, L\mathbf{u}\rangle + \lambda F(\mathbf{u})$ and a hard thresholding that minimizes the double-well potential term. More precisely, the updated solution $\mathbf{u}^{t+1}$ from the t-th iteration is given by

$$\begin{cases} \mathbf{u}^{t+1/2} = \mathbf{u}^t - dt\big(L\mathbf{u}^t + \lambda \nabla F(\mathbf{u}^t)\big) \\ \mathbf{u}^{t+1} = \mathcal{H}_{1/2}(\mathbf{u}^{t+1/2}), \end{cases} \tag{10}$$

where ∇F is the gradient of F, $dt > 0$ is a time stepsize, and $\mathcal{H}_{1/2}(\cdot)$ is a hard thresholding operator defined as

$$(\mathcal{H}_{1/2}(\mathbf{u}))_i = \begin{cases} 1, & \text{if } u_i \geq 1/2 \\ 0, & \text{if } u_i < 1/2, \end{cases} \tag{11}$$

for $i = 1, \ldots, n$. To circumvent the restriction on binary solutions in the MBO scheme, we use a bitwise scheme to deal with grayscale images in Section III.

III. Proposed Method

Let $X \in \mathbb{R}^{w\times n}$ be a hyperspectral image, where w is the number of spectral bands and n is the number of pixels in the image. We denote the spectral signature of pure materials, called *endmembers*, as $\{\mathbf{s}_j\}_{j=1}^k$ with k being the number of endmembers. Assume that the spectral signature at each pixel, namely each column of X, follows the standard linear mixing model, i.e.,

$$\mathbf{x}_i = \sum_{j=1}^k a_{ji}\mathbf{s}_j, \quad i = 1, \ldots, n, \tag{12}$$

where a_{ji} is the proportion of the j-th material at the i-th pixel. By concatenating all spectral signatures $\mathbf{s}_j$'s, we obtain a matrix $S \in \mathbb{R}^{w\times k}$, which is called the *mixing matrix*. Similarly, by assembling all weights a_{ji}'s, we obtain a matrix $A \in \mathbb{R}^{k\times n}$, which is called the *abundance map*. Thus we can rewrite (12) as $X = SA$. Different from [49], our method does not require the presence of pure pixels, rather just to assume the linear unmixing model (12).

By taking the noise into consideration, the blind unmixing problem is to estimate both S and A simultaneously from the noisy hyperspectral data X, i.e.,

$$X = SA + \eta,$$

where $\eta \in \mathbb{R}^{w\times n}$ is an additive noise term, which is typically assumed to be Gaussian noise. This is a highly ill-posed problem, and hence additional assumptions and regularizations are required. First, due to the physical interpretation of (12), both S and A are assumed to be nonnegative matrices, i.e., $S \in \Omega_{w\times k}$ and $A \in \Omega_{k\times n}$ with Ω defined in (2). In addition, since each element of A is the proportion of one of the pure materials in a single pixel, it is natural to impose the sum-to-one assumption, i.e., $\mathbf{1}_k^\top A = \mathbf{1}_n^\top$, where $\mathbf{1}_m$ denotes the all-one (column) vector of length m. We use the above two assumptions as constraints to refine the solution space.

In the previous work [50], we considered a graph Laplacian regularization for hyperspectral unmixing, i.e.,

$$J_{H_1}(A) = \frac{1}{2}\sum_{i,j=1}^n \|\widehat{\mathbf{a}}_i - \widehat{\mathbf{a}}_j\|_2^2 W_{ij}, \tag{13}$$

where $\mathbf{a}_i$ is the i-th column of A and $\widehat{\mathbf{a}}_i = \mathbf{a}_i/\sqrt{d_{ii}}$. However, the graph Laplacian regularization usually causes oversmoothing due to the presence of ℓ_2-norm in (13). To mitigate the oversmoothing artifacts, we propose a graph total variation (gTV) regularization on the abundance map, i.e.,

$$J_{TV}(A) = \frac{1}{2}\sum_{i,j=1}^n \|\widehat{\mathbf{a}}_i - \widehat{\mathbf{a}}_j\|_1 W_{ij}. \tag{14}$$

Minimizing J_{TV} can preserve edges of the abundance map for each material in a nonlocal fashion. The proposed gTV-regularized model for blind hyperspectral unmixing can be formulated as

$$\min_{\substack{S\in\Omega_{w\times k} \\ A\in\Omega_{k\times n}, \mathbf{1}_k^\top A = \mathbf{1}_n^\top}} \frac{1}{2}\|X - SA\|_{\mathrm{F}}^2 + \lambda J_{TV}(A), \tag{15}$$

where λ is a positive tuning parameter. Note that we use the given hyperspectral data X to generate a weighted graph by assuming that spectral signatures and abundance maps share the same spatial smoothness. Note that the sum-to-one constraint on the abundance map is commonly used in hyperspectral unmixing [2]; it implicitly enforces sparsity because it is related to the ℓ_1-norm. By considering only the sparsity of spatial gradients, the spatial TV regularization has a tendency to oversmooth the abundance map [51]. On the contrary, the proposed gTV regularization considers the similarity of spectral information at different pixels and hence it can preserve fine spatial features in the abundance map.

In order to apply the ADMM framework, we rewrite the constraints in (15) using indicator functions. In general, the indicator function χ_Δ of a set Δ is defined as

$$\chi_\Delta(Z) = \begin{cases} 0, & Z \in \Delta; \\ \infty, & \text{otherwise.} \end{cases}$$

By denoting $\Pi := \{Z \in \mathbb{R}^{k\times n} : Z \in \Omega_{k\times n}, \mathbf{1}_k^\top Z = \mathbf{1}_n^\top\}$, we can rewrite the model (15) as an unconstrained problem,

$$\min_{S,A} \frac{1}{2}\|X - SA\|_{\mathrm{F}}^2 + \lambda J_{TV}(A) + \chi_{\Omega_{w\times k}}(S) + \chi_\Pi(A). \tag{16}$$

We introduce two auxiliary variables $B \in \mathbb{R}^{k\times n}$, $C \in \mathbb{R}^{w\times k}$ and rewrite the objective function (16) as its equivalent form,

$$\begin{aligned} \min_{S,A,B,C} \frac{1}{2}\|X - CA\|_{\mathrm{F}}^2 + \lambda J_{TV}(B) + \chi_{\Omega_{w\times k}}(S) + \chi_\Pi(A) \\ \text{s.t.} \quad A = B,\ S = C. \end{aligned} \tag{17}$$

The corresponding augmented Lagrangian is

$$\mathcal{L} = \frac{1}{2}\|X - CA\|_{\mathrm{F}}^2 + \lambda J_{TV}(B) + \chi_{\Omega_{w\times k}}(S) + \chi_{\Pi}(A) + \frac{\rho}{2}\|A - B + \widetilde{B}\|_{\mathrm{F}}^2 + \frac{\gamma}{2}\|S - C + \widetilde{C}\|_{\mathrm{F}}^2,$$

where $\widetilde{B}$, $\widetilde{C}$ are dual variables and ρ, γ are two positive parameters. Then the ADMM algorithm requires solving four subproblems at each iteration, i.e., minimizing $\mathcal{L}$ with respect to C, S, A and B individually while fixing the others. Specifically, the C-subproblem reads as

$$\underset{C}{\operatorname{argmin}} \frac{1}{2}\|X - CA\|_{\mathrm{F}}^2 + \frac{\gamma}{2}\|S - C + \widetilde{C}\|_{\mathrm{F}}^2,$$

which has a closed-form solution. The S-subproblem seeks the projection of $C - \widetilde{C}$ onto the set of all nonnegative matrices, which can be solved by hard thresholding. As for the A-subproblem, the solution can be obtained by projecting a least squares solution onto the convex set Π, i.e.,

$$A = \mathcal{P}_{\Pi}\Big((S^\top S + \rho I)^{-1}\big(S^\top X + \rho(B - \widetilde{B})\big)\Big), \tag{18}$$

where $\mathcal{P}_{\Pi}$ is the projection operator on the set Π that can be implemented by a fast algorithm [52].

For the B-subproblem, we approximate the nondifferentiable gTV by the graph GL functional. To remove the binary restriction of MBO, we approximate any real number in $[0, 1]$ by its best M-bit binary representation [36]. We apply the MBO scheme on each channel separately, which can be implemented in parallel. Finally, we combine all the channels to get an approximated solution with elements in $[0, 1]$ for the B-subproblem. In all our experiments, we set $M = 8$. More specifically, we approximate the matrix B by a set of M binary matrices $B_m \in \mathbb{R}^{k\times n}$ with $m = 1, \cdots, M$ such that

$$B_{ij} \approx \sum_{m=1}^{M} 2^{-m}(B_m)_{ij}, \tag{19}$$

where M is the total number of bits being considered and B_m is the m-th bit channel of the matrix B, i.e., $(B_m)_{ij} \in \{0, 1\}$. Likewise, we approximate A and $\widetilde{B}$ in the same manner and get two sets of binary matrices $\{A_m\}_{m=1}^M$ and $\{\widetilde{B}_m\}_{m=1}^M$. Then for each channel, we approximate the gTV regularization J_{TV} by the graph GL functional (8). Note that $\langle A^\top, L_s A^\top\rangle = \operatorname{tr}(AL_sA^\top)$ due to (5) and hence we obtain the following minimization problem for each B_m,

$$\min_{B_m} \ \varepsilon \operatorname{tr}(B_m L_s B_m^\top) + \frac{1}{\varepsilon}\Phi(B_m) + \frac{\rho}{2\lambda}\|B_m - A_m - \widetilde{B}_m\|_{\mathrm{F}}^2, \tag{20}$$

where L_s and Φ are defined in Section II. Note that we assume that the graph structure at each channel is consistent with the one that is defined by the given hyperspectral data X.

We apply the MBO scheme (10) to minimize (20), which is a two-step iterative algorithm. In particular, the first step requires solving for $B_m^\top$ from

$$L_s B_m^\top + \frac{\rho}{\lambda}(B_m^\top - A_m^\top - \widetilde{B}_m^\top) = 0. \tag{21}$$

Motivated by [41], we further accelerate the MBO by taking advantage of the approximated eigendecomposition of L_s given in (7). Multiplying both sides of (21) with $V^\top$ from the left, we get $\Lambda V^\top B_m^\top + \frac{\rho}{\lambda}\Big(V^\top B_m^\top - V^\top(A_m^\top + \widetilde{B}_m^\top)\Big) = 0$, or equivalently

$$B_m V\Lambda + \frac{\rho}{\lambda}\Big(B_m V - (A_m + \widetilde{B}_m)V\Big) = 0, \tag{22}$$

since $V^\top V = I$. As a result, we only need to solve for $B_m V \in \mathbb{R}^{k\times p}$ with a reduced problem size. Denote $Z_m = B_m V$ and $D_m = \frac{\rho}{\lambda}\big(B_m V - (A_m + \widetilde{B}_m)V\big)$. At the $(\tau+1)$-th iteration, we have the following algorithm to update B_m:

$$\begin{cases} Z_m^{\tau+1} = Z_m^\tau(I - d\tau\Lambda) - d\tau \cdot D_m^\tau \\ B_m^{\tau+1/2} = V Z_m^{\tau+1} \\ D_m^{\tau+1} = \frac{\rho}{\lambda}\Big(B_m^{\tau+1/2} - (A_m + \widetilde{B}_m)\Big)V \\ B_m^{\tau+1} = \mathcal{H}_{1/2}(B_m^{\tau+1/2}). \end{cases} \tag{23}$$

Here the first three equations in (23) are obtained by applying fixed-point iteration to solve (22), and the last equation in (23) is from the MBO scheme in (10). Our numerical experiments show that five iterations of (23) for each B_m-subproblem are sufficient to produce reasonable results. If the B-subproblem can be solved within certain accuracy, then the convergence of ADMM can be guaranteed [53].

In summary, each subproblem in the ADMM algorithm can be solved efficiently either through a closed-form solution or within a few iterations. The entire algorithm is presented in Algorithm 1, which terminates when either the relative error between two subsequent mixing matrices, i.e., $\|S^t - S^{t+1}\|_{\mathrm{F}}/\|S^t\|_{\mathrm{F}}$, or the relative error between two subsequent abundance maps, i.e., $\|A^t - A^{t+1}\|_{\mathrm{F}}/\|A^t\|_{\mathrm{F}}$, is smaller than a given tolerance.

Algorithm 1 Blind Hyperspectral Image Unmixing Based on the Graph TV and MBO

Input: data X; parameters ρ, λ, maximum numbers of outer/inner loops T_{out}/T_{in}, and tolerance tol.
Output: S and A.
Initialize: S^0, A^0, and use the Nyström method to get the reduced eigendecomposition form of the graph Laplacian $L = V\Lambda V^\top$.
for $t = 0, \ldots, T_{out} - 1$ **do**
  $C^{t+1} = (X(A^t)^T + \gamma(S^t + \widetilde{C}^t))(A^t(A^t)^T + \gamma I)^{-1}$.
  $S^{t+1} = \max(C^{t+1} - \widetilde{C}^t, 0)$.
  $A^{t+1} = \mathcal{P}_{\Pi}\Big(((S^t)^T S^t + \rho I)^{-1}((S^t)^T X + \rho(B^t - \widetilde{B}^t))\Big)$.
  Bitwise update B^{t+1} via (23) with $\tau = 1, \ldots, T_{in}$.
  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

Here we discuss the complexity of the proposed algorithm and compare it with the other two related methods. The computational complexity of the Nyström method is $O(wpn + p^2 n)$, mainly for computing W_{12} and singular value decomposition in (6). This is much smaller than calculating the graph Laplacian matrix directly as described in Section II-A, which is $O(wn^2)$. As for the space complexity, using the

approximated graph Laplacian requires storing only $O(pn)$ numbers, while using the full graph Laplacian would need to store $O(n^2)$ numbers.

The time complexity of each step in Algorithm 1 is summarized as follows:

- C update: $O(wkn)$;
- S update: $O(wk)$;
- A update: $O(wkn + nk\log k) = O(wkn)$;
- B update per bit channel: $O(kpn)$;
- $\tilde{B}, \tilde{C}$ update: $O(kn)$.

Therefore, the time complexity for our algorithm per iteration is $O(kn(w+p))$ in total. Given $p \ll n$ and $k < w$, this is faster than the other two related methods: SUnSAL-TV [15] and GLNMF [31], which are in the order of $O(wn(w+\log n))$ and $O(kn(w+kn))$, respectively.

IV. Numerical Experiments

In this section, we conduct extensive experiments on synthetic and real data to demonstrate the performance of the proposed approach, referred to as "gtvMBO", in comparison with the state-of-the-art methods in blind and nonblind hyperspectral unmixing. Methods that we compare include FCLSU [11], SUnSAL-TV [15] (denoted by STV), GLNMF [31], fractional norm ℓ_q regularized unmixing method with $q = 0.1$ (denoted by FRAC) [13], NMF-QMV [21] (denoted by QMV), and our earlier unmixing work based on the graph Laplacian [50] (denoted by GraphL).

To quantitatively measure the performance, we adopt the following metrics to calculate the error between an estimation $\widehat{Y} \in \mathbb{R}^{r\times c}$ and the reference $Y \in \mathbb{R}^{r\times c}$.

1) 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.

2) Normalized mean-square error (nMSE)

$$nMSE(Y,\widehat{Y}) = \frac{\|Y-\widehat{Y}\|_{\mathrm{F}}}{\|Y\|_{\mathrm{F}}}.$$

3) Spectral angle mapper (SAM) in degrees

$$SAM(Y,\widehat{Y}) = \frac{1}{c}\sum_{j=1}^{c}\arccos\left(\frac{\mathbf{y}_j^\top \hat{\mathbf{y}}_j}{\|\mathbf{y}_j\|_2\|\hat{\mathbf{y}}_j\|_2}\right),$$

where $\mathbf{y}_j \in \mathbb{R}^r$ is the j-th column of Y. The index j is skipped in the sum when $\|\mathbf{y}_j\|_2\|\hat{\mathbf{y}}_j\|_2 = 0$.

In order to make a fair comparison, we use the initialization steps in [13] for all the methods considered in this paper. In particular, VCA [54], which returns $10k$ endmember candidates that are clustered into k 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 an initial guess of S^0 and A^0, respectively, for all compared methods. We set $\sigma = 5$ in the weight computation (3) and randomly select 0.1% samples from the entire pixel list in the Nyström method to approximate the graph Laplacian. As for γ, ρ and λ, we choose the optimal parameters that minimize nMSE$(A, \widehat{A})$. We first perform a coarse grid search with parameter candidates evenly spaced over the interval on a log scale, then do a finer grid search around the best parameters, e.g., search for an optimal λ in $\{10^{2.5}, 10^{2.75}, \ldots, 10^{3.5}\}$ given $\lambda = 10^3$ from the coarse grid search. For GraphL and gtvMBO, the coarse grid search is over $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^5\}$, $\rho/\lambda \in \{10^{-3}, 10^{-2}, \ldots, 10^3\}$, and $\gamma \in \{10^2, 10^3, \ldots, 10^5\}$. For FRAC, we fix $\rho = 10$ as suggested in [13] and search for λ among $\{10^{-5}, 10^{-4}, \ldots, 10^5\}$. For QMV, we search for λ (denoted by β in [21]) $\in \{10^{-5}, 10^{-4}, \ldots, 10^5\}$. For GLNMF and STV, we search for $\lambda, \mu \in \{10^{-5}, 10^{-4}, \ldots, 10^5\}$. See Section V for a detailed discussion on parameter selection and sensitivity of our method. Our Matlab source codes are available at https://github.com/HarlinLee/gtvMBO-public. 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.

A. Synthetic Data

To evaluate the performance of all methods, we construct a set of synthetic data X with ground truth mixing matrix S and endmember matrix A. Fig. 1 shows the ground truth abundance maps. We adopt the same simulation procedure as in [15], where an endmember library is generated by randomly selecting 240 materials from the USGS 1995 library with 224 spectral bands. The noise-free hyperspectral image with 75×75 pixels is generated by a random selection of 5 spectral signatures from the library. The respective ground truth abundances are randomly fixed as 0.1149, 0.0741, 0.2003, 0.2055, and 0.4051. The noisy hyperspectral data is then obtained by adding zero-mean Gaussian noise with a signal-to-noise ratio (SNR) of 10db and 20db, respectively.

Table. I compares all methods on the noisy data quantitatively. To get a visual comparison, we present the case of SNR$=$ 10dB in Fig. 2. In particular, we show all the reconstructed abundance maps corresponding to the fifth ground-truth abundance in Fig. 1. We exclude the results of FCLSU and FRAC in Fig. 2, as both fail to recover the abundance maps under such a low SNR scenario. One can see that STV and GLNMF have a different color range on the background comparing to other methods, while the QMV background is still noisy. The proposed gtvMBO achieves a balance between recognizable objects and background noise, while the result of GraphL is slightly oversmoothed. Note that the proposed gtvMBO only considers the regularization on A, while QMV uses the minimum-volume based regularization on S, but our method still gives comparable results in recovering S compared to QMV, and has an advantage on reconstructing A, especially when the underlying abundance map has spectral geometries. In addition, gtvMBO can reconstruct A well within a few iterations but it takes more iterations to get a good reconstruction of S. In the preprocessing step, both GraphL and gtvMBO take less than a second to estimate the eigenvalues and eigenvectors of the low-rank approximation to the graph Laplacian by the Nyström method, while GLNMF

	FCLSU	FRAC	STV	GLNMF	QMV	GraphL	gtvMBO
				SNR $= 10$			
RMSE$(A, \hat{A})$	0.242	0.157	0.248	0.24	0.093	0.0513	**0.051**
nMSE$(A, \hat{A})$	1.05	0.696	1.07	1.03	0.435	0.364	**0.327**
RMSE$(S, \hat{S})$	**0.14**	–	–	0.211	0.612	0.16	0.16
nMSE$(S, \hat{S})$	**0.205**	–	–	0.321	0.881	0.244	0.241
SAM$(S, \hat{S})$	10.2	–	–	14.8	40.5	8.65	**8.57**
				SNR$= 20$			
RMSE$(A, \hat{A})$	0.106	0.106	0.065	0.107	0.048	**0.043**	0.065
nMSE$(A, \hat{A})$	0.523	0.523	0.314	0.521	**0.227**	0.242	0.314
RMSE$(S, \hat{S})$	0.055	–	–	0.067	**0.037**	0.095	0.054
nMSE$(S, \hat{S})$	0.092	–	–	0.104	**0.053**	0.13	0.091
SAM$(S, \hat{S})$	2.67	–	–	3.18	2.69	6.88	**2.65**

TABLE I: Unmixing results on the synthetic dataset.

	FCLSU	FRAC	STV	GLNMF	QMV	GraphL	gtvMBO
RMSE$(S, \hat{S})$	0.044	–	–	**0.036**	0.073	0.052	0.070
nMSE$(S, \hat{S})$	0.169	–	–	**0.153**	0.302	0.203	0.296
SAM$(S, \hat{S})$	**3.64**	–	–	4.49	12.8	7.86	9.84
RMSE$(A, \hat{A})$	0.18	0.165	0.165	0.187	0.148	0.139	**0.096**
nMSE$(A, \hat{A})$	0.455	0.429	0.375	0.502	0.428	0.302	**0.243**
Graph time (sec)	–	–	–	66.4	–	0.082	0.082
Alg. time (sec)	2.34	0.052	4.08	8.73	1.6	0.094	0.609
λ	–	1	0.01	1	$10^{2.75}$	$10^{-5.25}$	$10^{-3.75}$
ρ	–	10	–	–	–	$10^{-1.75}$	$10^{-2.25}$
γ	–	–	–	–	–	10^{5}	10^{4}
μ	–	–	–	1	–	–	–
Iterations	–	2	1000	1000	101	30	30

TABLE II: Unmixing results on the Samson dataset.

typically takes a minute to calculate the graph Laplacian. In terms of running time, gtvMBO is slower than FRAC and GraphL, but much faster than the other competing methods.

B. Real Data

We use the real hyperspectral data X with the references S and A from [55], including Samson, Jasper Ridge and Urban data sets. In particular, the endmembers are manually selected from the image data by assuming k distinct materials with one signature per material and neglecting possible spectral variability issues. The reference abundances are obtained via FCLSU. This way of generating references for endmembers/abundances has been widely used for assessing the performance of various unmixing algorithms. As no ground-truth is available for the real data, it is common to compare the unmixing results to the reference endmembers/abundances.

1) Samson: In the first experiment, we use the Samson data with 95×95 pixels and 156 spectral bands after preprocessing, whose reference has three endmembers. The unmixing results are given in Figs. 3-4 and Table II for endmembers, abundance maps, and quantitative metrics, respectively. In Fig. 3, all endmember plots can capture the rough shape and discontinuities in the ground truth but with different heights. The gtvMBO result has many endmember elements that are close to zero since we enforce the nonnegative constraint on the endmember S by using the hard thresholding operator in the S-subproblem. For the abundance maps, the STV results look blurry when trying to preserve spatial smoothness and the GLNMF results are noisy in the homogeneous areas, as its graph Laplacian is based on the entire data that may contain certain amount of noise. Both blurring and noisy artifacts can be mitigated by the low-rank approximation of graph Laplacian in the Nyström method as in GraphL and gtvMBO. On the other hand, gtvMBO yields sharper edges than GraphL, thanks to the graph TV regularization. Table II reports that GLNMF gives the best estimations in S at the cost of high computational costs, whereas the proposed method is the best in reconstructing the abundance maps. Note that "graph time" in Table II is referred to as the time needed to compute the adjacency matrix (for GLNMF) and the graph Laplacian matrix (for GraphL and gtvMBO), while "algorithm time," or "alg. time" in short, refers to the time needed to run the unmixing algorithm after initialization and graph construction. The overall computation time of gtvMBO is the sum of "graph time" and "time," which is comparable to QMV and much faster than GLNMF.

2) Jasper Ridge: In the second experiment, we test the Jasper Ridge data which has 100×100 pixels and 198 spectral bands. The unmixing results for endmembers and abundance maps are shown in Figs. 5-6. In Fig. 6, the FRAC abundance maps have the highest image contrast, while mistakenly identifying trees and roads in some areas, especially the top right part. The STV abundance maps are over-smoothed, especially in the Dirt abundance map. Since only the five nearest neighbors are considered when calculating the pairwise weight of a fully-connected graph, GLNMF may miss some global features while preserving fine details. For example, some variations in the water are captured but some roads are not identified in the GLNMF abundance maps. One can see that both GraphL and gtvMBO perform very well at identifying Water and Road abundance maps because of the learned graph structure in the Nyström method. Specifically for the road abundance, these two methods can recover the road on the rightmost part of the image. This phenomenon could be explained by the fact that it is a very narrow structure and the nonlocal similarity with road pixels across all bands plays an important role, illustrating an advantage of using graph TV over spatial TV. The gtvMBO results are even better than GraphL in preserving the sharpness especially in the Dirt abundance map. The endmember spectral plot in Fig. 5 also confirms that the methods failing for the road extract a very poor signatures compared to the reference. Table III compares all the methods quantitatively. It is true that QMV gives the best results on this dataset, which is probably because that the assumptions made by QMV hold on Jasper, but not on the other data sets. The proposed gtvMBO can recover endmembers and abundance maps in a balanced manner. The comparison results imply that a good RMSE on the reconstructed data can not guarantee a good unmixing

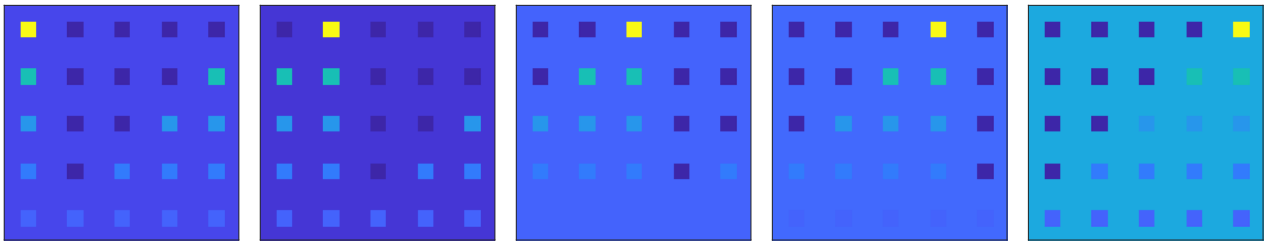

Fig. 1: Ground truth abundance maps of the synthetic data (five endmembers).

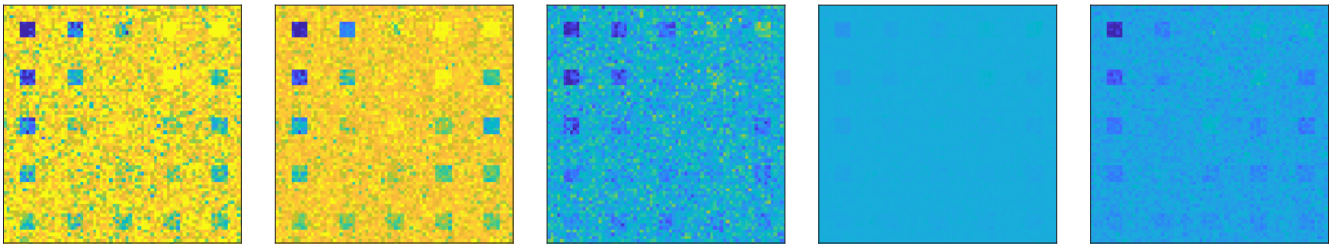

Fig. 2: Reconstructed abundance maps of the fifth element from the noisy data with SNR 10dB. All images are visualized over the range $[0, 1]$.

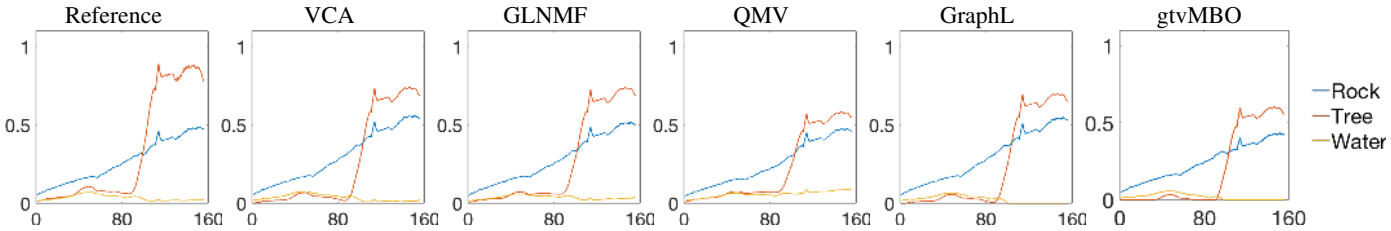

Fig. 3: Endmember profiles (S) of the Samson dataset.

	FCLSU	FRAC	STV	GLNMF	QMV	GraphL	gtvMBO
RMSE$(S, \hat{S})$	0.144	–	–	0.133	**0.031**	0.18	0.083
nMSE$(S, \hat{S})$	0.608	–	–	0.598	**0.107**	0.629	0.288
SAM$(S, \hat{S})$	16.8	–	–	14.9	**3.54**	14.6	12.8
RMSE$(A, \hat{A})$	0.148	0.109	0.142	0.111	**0.073**	0.145	0.136
nMSE$(A, \hat{A})$	0.472	0.46	0.47	0.437	**0.221**	0.38	0.353
Graph time (sec)	–	–	–	126	–	0.225	0.225
Alg. time (sec)	4.27	9.52	4.56	10.4	3.89	0.34	3.32
λ	–	1	$10^{-1.25}$	$10^{-0.5}$	$10^{2.25}$	$10^{-4.5}$	$10^{-4.25}$
ρ	–	10	–	–	–	0.1	$10^{-2.75}$
γ	–	–	–	–	–	10^4	$10^{3.75}$
μ	–	–	–	$10^{-2.5}$	–	–	–
Iterations	–	300	1000	1000	101	100	100

TABLE III: Unmixing results on the Jasper Ridge dataset.

performance.

3) Urban: Lastly, we test a relatively large dataset - the Urban dataset with 307×307 pixels and 162 spectral bands, whose reference has four endmembers. The results for all methods are presented in Figs. 7-8. In Fig. 8, most methods, including FCLSU, FRAC, STV, GLNMF, and QMV, yield abundance maps in low image contrast due to the initial guess, especially in the abundance maps for the asphalt and roof. As a by-product, the proposed gtvMBO can greatly improve the image contrast of the abundance map due to the graph TV regularization. In addition, all the methods have a hard time extracting a good roof endmember, but the graph-based approaches are able to compensate this with more features preserved. Also note that because QMV does not enforce non-negativity on S, the resulting spectrum for Asphalt in QMV goes below zero. In the Roof abundance maps, only GraphL and gtvMBO can capture those sporadic roof tops since the approximated graph Laplacian considers the pairwise similarity across spectral bands in the original data with dimension w much greater than the dimension k for the column space of the abundance map A. In Table IV, we list all quantitative metric comparisons where gtvMBO reaches the smallest residual error and get comparable reconstruction errors for the abundance map and endmember with GraphL. Overall, the proposed method can reconstruct abundance maps and endmember matrices with high accuracy in a short time.

V. Discussion

In this section, we discuss parameter selection in our algorithm. Due to heavy computations involved in these tasks, all the results presented in this section are performed on a

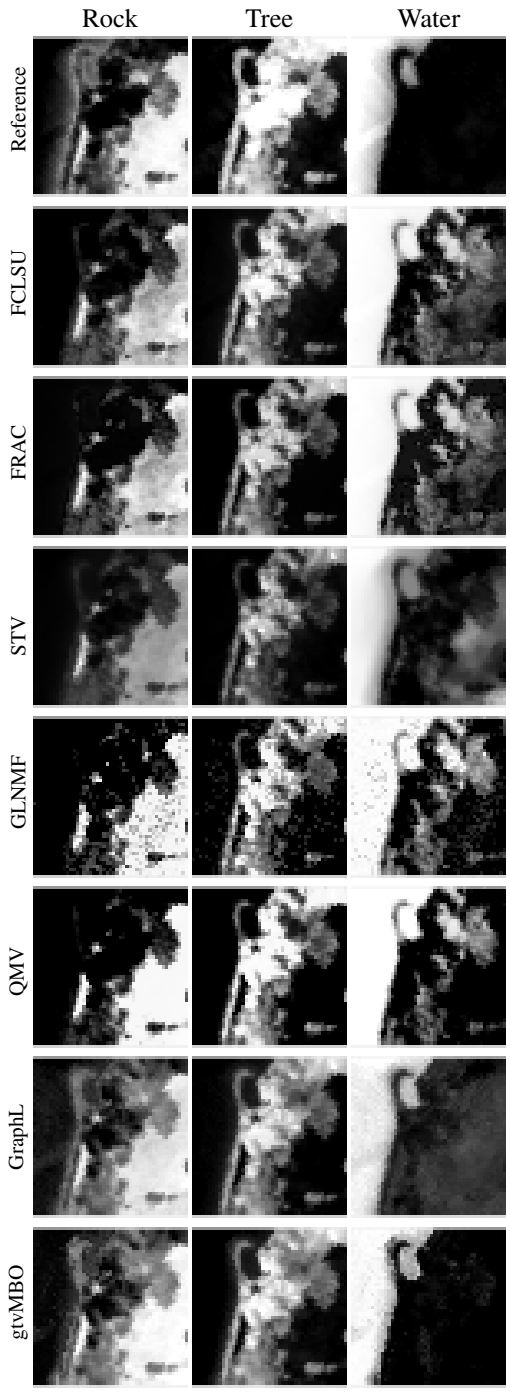

Fig. 4: Abundance maps (A) of the Samson dataset.

workstation of DELL R7425 Dual Processor AMD Epyc 32 core 2.2 GHz machines with 512GB RAM each.

There are several tuning parameters in our approach: the filtering parameter σ in computing pairwise weights of the graph, the regularization parameter λ associated with the graph TV in the proposed unmixing model, the penalty parameters ρ and γ in the proposed algorithm based on ADMM, and time step size dt for the diffusion step in the modified MBO scheme. The value of σ could be changed proportionally according to the number of spectral bands w. Since all the test datasets have 100$\sim$200 spectral bands, we find that $\sigma = 5$ typically gives good results, so we fix it throughout the experimental section. To solve the B-subproblem, we fix the step size $dt = 0.01$ and run 5 iterations of (23) in the modified MBO scheme.

	FCLSU	FRAC	STV	GLNMF	QMV	GraphL	gtvMBO
RMSE$(S, \hat{S})$	0.109	–	–	0.188	0.211	**0.099**	**0.099**
nMSE$(S, \hat{S})$	**0.635**	–	–	1.35	1.2	0.636	0.639
SAM$(S, \hat{S})$	19.5	–	–	17.9	46.4	**14.8**	14.9
RMSE$(A, \hat{A})$	0.145	0.153	0.289	0.175	0.245	**0.134**	0.136
nMSE$(A, \hat{A})$	0.438	0.45	0.756	0.554	0.655	**0.384**	0.393
Graph time (sec)	–	–	–	10800	–	9.09	9.09
Alg. time (sec)	34.7	0.85	142	86.1	29	0.353	22.4
λ	–	$10^{-0.5}$	$10^{-2.25}$	$10^{-1.5}$	$10^{1.75}$	$10^{-3.25}$	10^{-6}
ρ	–	10	–	–	–	$10^{-1.25}$	$10^{-5.5}$
γ	–	–	–	–	–	$10^{4.75}$	$10^{4.75}$
μ	–	–	–	$10^{-5.5}$	–	–	–
Iterations	–	2	1000	1000	101	10	10

TABLE IV: Unmixing results on the Urban dataset.

To find optimal or sub-optimal values of λ, ρ, and γ, we consider a skillful strategy which alleviates the time-consuming parameter tuning. If the value of λ increases, the recovered abundance map A has a graph structure more similar to that of the given data X but with larger residual error and vice versa. The penalty parameters ρ and γ both control the convergence of the proposed algorithm according to the ADMM framework. In other words, λ is a model parameter that affects the performance and ρ, γ are algorithmic parameters that affect the convergence. Therefore, we suggest a set of default parameters by fixing the ratios as $\rho/\lambda = 1$, $\gamma/\lambda = 10^7$ and only tuning the regularization parameter λ. In fact, the B-subproblem is determined by the ratio ρ/λ. Table V shows that using these default algorithmic parameters still ensures comparable unmixing performance on the datasets to when we tune all the three parameters together. Note that the optimal parameters indeed yield better results than the default parameters in terms of SAM$(S, \hat{S})$, which is due to the fact that our regularization is formulated on A and the optimal parameters are determined according to nMSE$(A, \hat{A})$, resulting in more deviations in S. In future work, we might consider choosing optimal parameters based on a combination of evaluation metrics on S and A.

In addition, learning a graph Laplacian or its low-rank approximation is an important preprocessing step in our proposed method. In the Nyström method, the sampling rate is fixed as 0.1% in all our experiments. Our empirical results show that this is sufficient for preserving the graph structure of the original hyperspectral data. In fact, there is a trade-off between the number of samples corresponding to the rank of the approximated Laplacian and the orthogonality of columns in the approximated eigenvectors: more samples can improve accuracy in approximating the graph Laplacian but may result in loss of orthogonality of the resulting eigenvectors, which is also desired in our modified MBO scheme (23). Other adaptive sampling schemes for the Nyström extension [56] will be

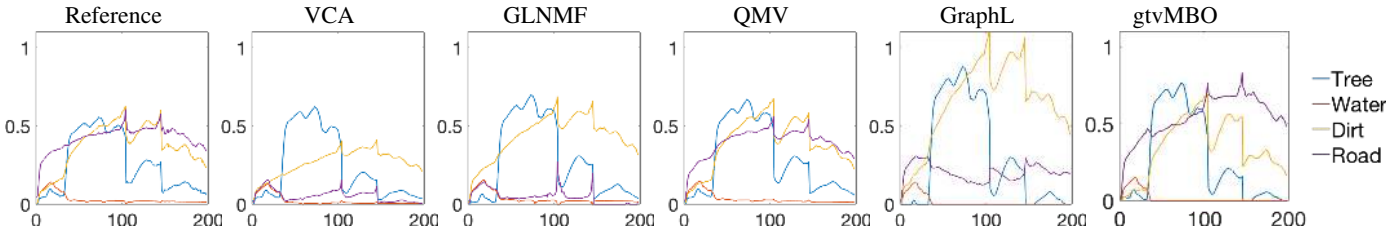

Fig. 5: Endmember profiles (S) of the Jasper Ridge dataset.

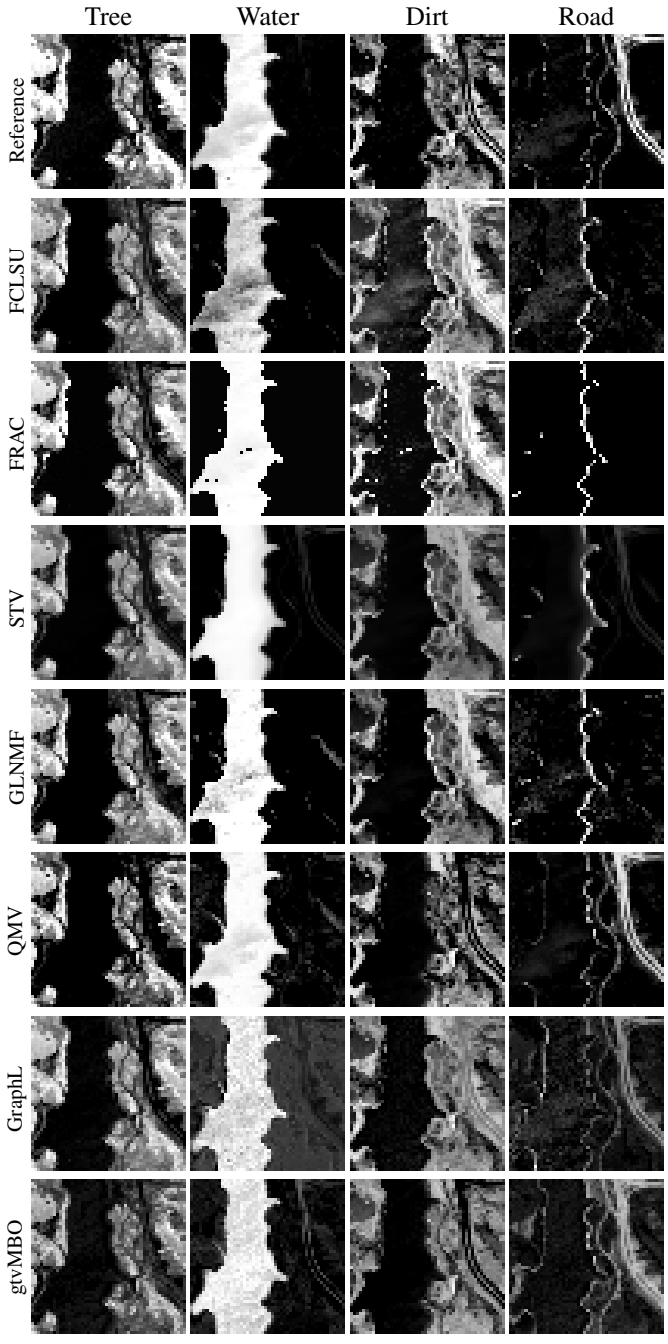

Fig. 6: Abundance maps (A) of the Jasper Ridge dataset.

	Samson	**Jasper**	**Urban**
RMSE($S, \hat{S}$)	0.07 / **0.062**	**0.083** / 0.13	**0.099** / 0.10
nMSE($S, \hat{S}$)	0.3 / **0.23**	**0.29** / 0.44	**0.64** / 0.67
SAM($S, \hat{S}$)	**9.84** / 16.1	**12.8** / 17.8	**14.9** / 15.8
RMSE($A, \hat{A}$)	**0.096** / 0.12	0.14 / **0.13**	0.14 / 0.2
nMSE($A, \hat{A}$)	**0.24** / 0.27	0.35 / 0.35	**0.39** / 0.41
λ	$10^{-3.5}$	10^{-8}	$10^{-2.5}$

TABLE V: Unmixing results of gtvMBO in A/B format, where A is the previous result using optimally tuned λ, ρ, γ, and B is the result of using default ratios $\rho/\lambda, \gamma/\lambda$ and only tuning the λ value (given in the last row.)

explored in our future work. For high performance computing applications, the Nyström loop can be optimized for specific architectures as in [57].

VI. Conclusions

We propose a graph TV regularized approach for blind hyperspectral unmixing to estimate both the abundance map and the mixing matrix under the assumption that the underlying abundance map and the given hyperspectral data share the same graph structure. In particular, we applied the Nyström method to approximate the eigenvalues and eigenvectors of a normalized graph Laplacian. To solve the proposed gTV regularized unmixing problem with probability simplex constraints, we derived an efficient algorithm based on ADMM. One of the subproblems is decomposed into bits and then solved by the fast MBO scheme at each bit channel. Extensive experiments were conducted to demonstrate that the proposed framework is effective and efficient, especially when the hyperspectral data have similarities across spectral bands. In the future, one could integrate robust graph learning methods and minimum-volume based regularizations into hyperspectral unmixing.

Acknowledgements

The initial research for this effort was conducted at the Research Collaboration Workshop for Women in Data Science and Mathematics, July 29-August 2, 2019, held at ICERM. Funding for the workshop was provided by ICERM. We would like to acknowledge Linda Ness for her helpful suggestions and discussions.

References

[1] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, 2000.

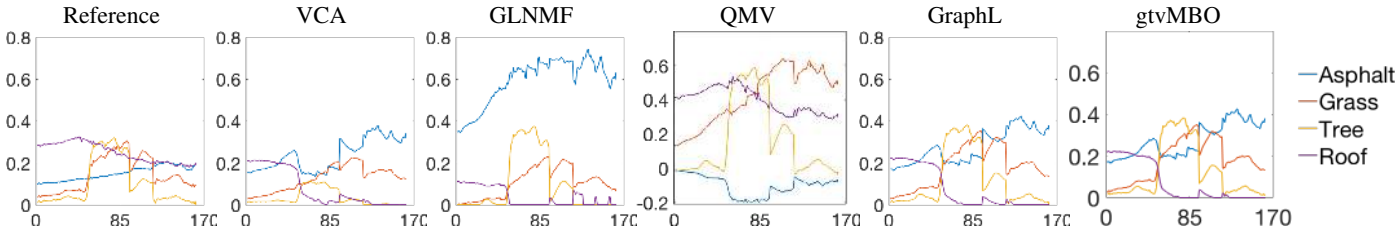

Fig. 7: Endmember profiles (S) of the Urban dataset.

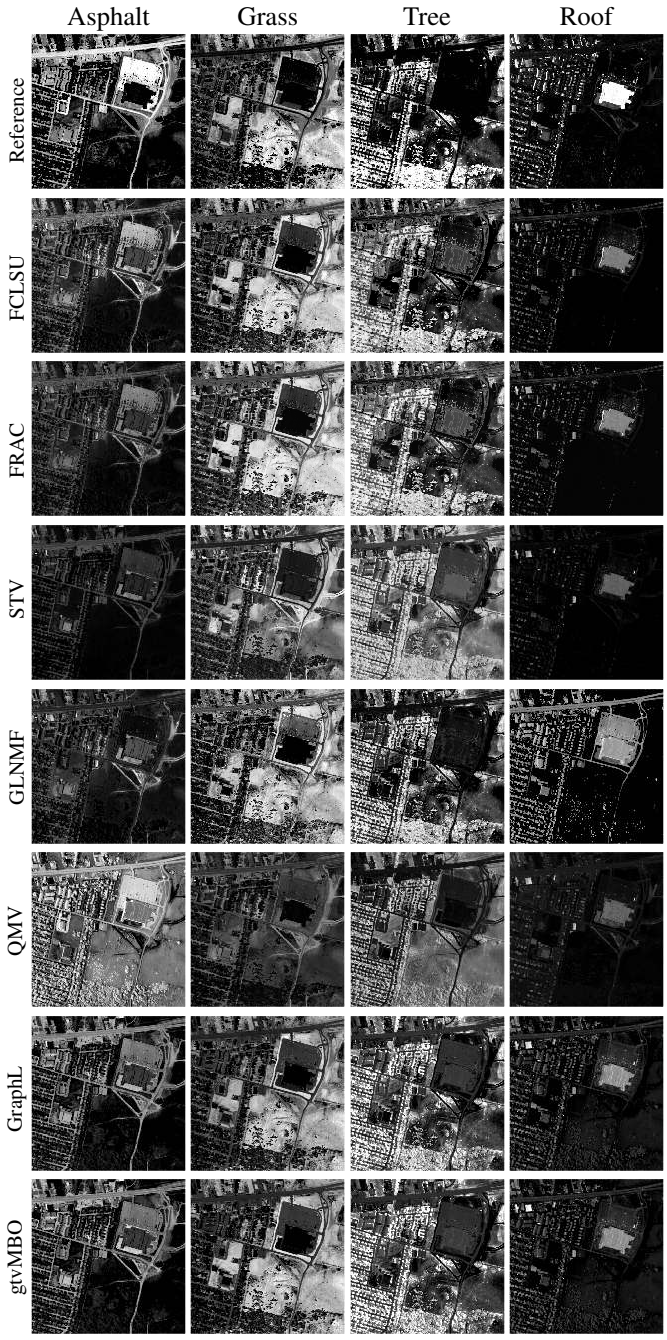

Fig. 8: Abundance maps (A) of the Urban dataset.

[2] 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.

[3] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Obser. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, 2014.

[4] L. Drumetz, M. A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.

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

[6] P. Sajda, S. Du, T. Brown, L. Parra, and R. Stoyanova, "Recovery of constituent spectra in 3D chemical shift imaging using nonnegative matrix factorization," in *4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pp. 71–76, 2003.

[7] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.

[8] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[9] 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.

[10] 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.

[11] 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.

[12] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, 2011.

[13] 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.

[14] L. I. Rudin, S. J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[15] 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.

[16] 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*, vol. 7334, p. 73341M, International Society for Optics and Photonics, 2009.

[17] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, 2017.

[18] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, 2018.

[19] Y. Yuan, Z. Zhang, and Q. Wang, "Improved collaborative non-negative matrix factorization and total variation for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 998–1010, 2020.

[20] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, 2016.

[21] L. Zhuang, C.-H. Lin, M. A. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9858–9877, 2019.

[22] G. Gilboa and S. J. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model Sim.*, vol. 7, no. 3, pp. 1005–1028, 2008.

[23] Y. Lou, X. Zhang, S. J. Osher, and A. L. Bertozzi, "Image recovery via nonlocal operators," *J. Sci. Comput.*, vol. 42, no. 2, pp. 185–197, 2010.

[24] W. Zhu, V. Chayes, A. Tiard, S. Sanchez, D. Dahlberg, A. L. Bertozzi, S. J. Osher, D. Zosso, and D. Kuang, "Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2786–2798, 2017.

[25] J. Yao, D. Meng, Q. Zhao, W. Cao, and Z. Xu, "Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2991–3006, 2019.

[26] K. Benzi, V. Kalofolias, X. Bresson, and P. Vandergheynst, "Song recommendation with non-negative matrix factorization and graph total variation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2439–2443, 2016.

[27] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[28] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, 2014.

[29] 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.

[30] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS J Photogramm Remote Sens*, vol. 88, pp. 101–118, 2014.

[31] 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.

[32] M. Li, F. Zhu, A. J. X. Guo, and J. Chen, "A graph regularized multilinear mixing model for nonlinear hyperspectral unmixing," *Remote Sens.*, vol. 11, no. 19, p. 2188, 2019.

[33] 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.

[34] R. Ammanouil, A. Ferrari, and C. Richard, "Hyperspectral data unmixing with graph-based regularization," in *7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, June 2015.

[35] Y. Van-Gennip and A. L. Bertozzi, "Gamma-convergence of graph Ginzburg-Landau functionals," *Adv. Differential Equ.*, vol. 17, no. 11/12, pp. 1115–1180, 2012.

[36] J. A. Dobrosotskaya and A. L. Bertozzi, "A wavelet-Laplace variational technique for image deconvolution and inpainting," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 657–663, 2008.

[37] R. Glowinski and A. Marrocco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par, penalisation-dualité, d'une classe de problems de Dirichlet non lineares," *Revue Française d'Automatique, Informatique, et Recherche Opéationelle*, vol. 9, pp. 41–76, 1975.

[38] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Comp. Math. Appl.*, vol. 2, pp. 17–40, 1976.

[39] B. Merriman, J. K. Bence, and S. J. Osher, "Motion of multiple junctions: A level set approach," *J. Comput. Phys.*, vol. 112, no. 2, pp. 334–363, 1994.

[40] 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*, pp. 689–693, 2014.

[41] 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.

[42] 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*, pp. 209–222, Springer, 2015.

[43] A. L. Bertozzi and A. Flenner, "Diffuse interface models on graphs for classification of high dimensional data," *SIAM Review*, vol. 58, no. 2, pp. 293–328, 2016.

[44] S. Esedog and Y. H. R. Tsai, "Threshold dynamics for the piecewise constant Mumford–Shah functional," *J. Comput. Phys.*, vol. 211, no. 1, pp. 367–384, 2006.

[45] R. V. Kohn and P. Sternberg, "Local minimisers and singular perturbations," *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, vol. 111, no. 1-2, pp. 69–84, 1989.

[46] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus, "Multiclass data segmentation using diffuse interface methods on graphs," *IEEE Trans Pattern Anal. Mach. Intell*, vol. 36, no. 8, pp. 1600–1613, 2014.

[47] G. Iyer, J. Chanussot, and A. L. Bertozzi, "A graph-based approach for feature extraction and segmentation of multimodal images," in *IEEE International Conference on Image Processing*, pp. 3320–3324, 2017.

[48] G. Iyer, J. Chanussot, and A. L. Bertozzi, "A graph-based approach for data extraction and segmentation of multimodal images," *(to appear) IEEE Trans. Geosci. Remote Sens.*, 2019.

[49] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, 2012.

[50] J. Qin, H. Lee, J. Chi, J. Chanussot, Y. Lou, and A. L. Bertozzi, "Fast Blind Hyperspectral Unmixing based on Graph Laplacian," in *10th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5, 2019.

[51] X. Zhang, C. Li, J. Zhang, Q. Chen, J. Feng, L. Jiao, and H. Zhou, "Hyperspectral unmixing via low-rank representation with space consistency constraint and spectral library pruning," *Remote Sensing*, vol. 10, no. 2, p. 339, 2018.

[52] 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.

[53] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Prog.*, vol. 55, no. 1-3, pp. 293—-318, 1992.

[54] 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.

[55] F. Zhu, "Hyperspectral unmixing: Ground truth labeling, datasets, benchmark performances and survey," *arXiv preprint arXiv:1708.05125*, 2017.

[56] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling methods for the Nyström method," *J. Mach. Learn Res.*, vol. 13, no. Apr, pp. 981–1006, 2012.

[57] Z. Meng, A. Koniges, Y. H. He, S. Williams, T. Kurth, B. Cook, J. Deslippe, and A. L. Bertozzi, "OpenMP Parallelization and Optimization of Graph-based Machine Learning Algorithms," in *Proc. IWOMP 2016. Lecture Notes in Computer Science* (M. N. de Supinksi B. and M. Wahib, eds.), vol. 9930, pp. 17–31, Springer, 2016.

Jing Qin (M'17) is an assistant professor in the Department of Mathematics, University of Kentucky at Lexington, KY. She received her B.S. degree in mathematics from Xuzhou Normal University, Jiangsu, China, in 2005, her M.S. degree in mathematics from East China Normal University, Shanghai, China, in 2008, and her Ph.D. in applied mathematics from Case Western Reserve University, Cleveland, OH, in 2013. From 2013 to 2016, she was an assistant adjunct professor in the Department of Mathematics at the University of California at Los Angeles. From 2016 to 2019, she worked as an assistant professor in the Department of Mathematical Sciences, Montana State University, Bozeman, MT. Her research interests include mathematical image processing, numerical optimization and high-dimensional data analysis.

Harlin Lee received the B.S. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2016 and 2017, respectively. She is currently working toward the Ph.D. degree in electrical and computer engineering with Carnegie Mellon University, Pittsburgh, PA, USA. Her research interests include graph regularization, unsupervised learning, and clinical data analysis.

Jocelyn T. Chi is a Ph.D. student in the Statistics Department at North Carolina State University. She is advised by Ilse Ipsen, and her research interests lie at the intersection of randomized numerical linear algebra and statistical learning methods for large data.

Lucas Drumetz (M'17) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 2013, and the Ph.D. degree from the Université de Grenoble Alpes, Grenoble, in 2016. In 2017, he was a Visiting Assistant Professor with the University of California at Los Angeles (UCLA), Los Angeles, CA, USA. In 2017, he was a Visiting Researcher with the University of Tokyo, Tokyo, Japan. Since 2018, he has been an Associate Professor with the Department of Signal and Communications, IMT-Atlantique, Brest, France. His research interests include signal and image processing, inverse problems, optimization techniques and machine learning for remote sensing data. Dr. Drumetz received the 2017 Best Ph.D. Award from the Université de Grenoble Alpes.

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998. Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning and artificial intelligence. He has been a visiting scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he is an Adjunct Professor of the University of Iceland. In 2015-2017, he was a visiting professor at the University of California, Los Angeles (UCLA). He holds the AXA chair in remote sensing and is an Adjunct professor at the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing. Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in chareg of meetings and symposia (2017-2019). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Transactions on Image Processing and the Proceedings of the IEEE. He was the Editor-in-Chief of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2011-2015). In 2014 he served as a Guest Editor for the IEEE Signal Processing Magazine. He is a Fellow of the IEEE, a member of the Institut Universitaire de France (2012-2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, 2018-2019).

Yifei Lou has been an Assistant Professor in the Mathematical Sciences Department, University of Texas Dallas, since 2014. She received her Ph.D. in Applied Math from the University of California Los Angeles (UCLA) in 2010. After graduation, she was a postdoctoral fellow at the School of Electrical and Computer Engineering Georgia Institute of Technology, followed by another postdoc training at the Department of Mathematics, University of California Irvine from 2012-2014. Dr. Lou received the National Science Foundation CAREER Award in 2019. Her research interests include compressive sensing and its applications, image analysis (medical imaging, hyperspectral, imaging through turbulence), and (nonconvex) optimization algorithms.

Andrea L. Bertozzi (M'14) received the B.A., M.A., and Ph.D. degrees in mathematics from Princeton University, Princeton, NJ, USA, in 1987, 1988, and 1991, respectively. She was on the faculty of the University of Chicago, Chicago, IL, USA, from 1991 to 1995, and Duke University, Durham, NC, USA, from 1995 to 2004. From 1995 to 1996, she was the Maria Goeppert-Mayer Distinguished Scholar at the Argonne National Laboratory, Lemont, IL, USA. Since 2003, she has been a Professor of mathematics with the University of California at Los Angeles, Los Angeles, CA, USA, where she is currently the Director of Applied Mathematics. In 2012, she was appointed the Betsy Wood Knapp Chair of Innovation and Creativity. Her research interests include graphical models for machine learning, image inpainting, image segmentation, cooperative control of robotic vehicles, swarming, and fluid interfaces, and crime modeling. Dr. Bertozzi is a fellow of the Society for Industrial and Applied Mathematics, the American Mathematical Society, and the American Physical Society. She was a recipient of the SIAM Kovalevsky Prize in 2009 and the SIAM Kleinman Prize in 2019. She is a member of the US National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences. She has served as a Plenary/Distinguished Lecturer for both SIAM and AMS and has served as Associate Editor of the SIAM journals on Multiscale Modelling and Simulation, Mathematical Analysis, SIAM Review, and Image Processing.