Pointwise Mutual Information based Graph Laplacian Regularized Sparse Unmixing

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-SA 4.0

SUBMISSION DATE / POSTED DATE

19-10-2021 / 21-10-2021

CITATION

Kucuk, Sefa; Yuksel, Seniha Esen (2021): Pointwise Mutual Information based Graph Laplacian Regularized Sparse Unmixing. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16831330.v1>

DOI

[10.36227/techrxiv.16831330.v1](https://doi.org/10.36227/techrxiv.16831330.v1)

Pointwise Mutual Information based Graph Laplacian Regularized Sparse Unmixing

Sefa Kucuk, Seniha Esen Yuksel, *Member, IEEE*

Abstract—Sparse unmixing (SU) aims to express the observed image signatures as a linear combination of pure spectra known a priori and has become a very popular technique with promising results in analyzing hyperspectral images (HSI) over the past ten years. In SU, utilizing the spatial-contextual information allows for more realistic abundance estimation. To make full use of the spatial-spectral information, in this letter, we propose a pointwise mutual information (PMI) based graph Laplacian regularization for SU. Specifically, we construct the affinity matrices via PMI by modeling the association between neighboring image features through a statistical framework, and then we use them in the graph Laplacian regularizer. We also adopt a double reweighted ℓ_1 norm minimization scheme to promote the sparsity of fractional abundances. Experimental results on simulated and real data sets prove the effectiveness of the proposed method and its superiority over competing algorithms in the literature.

Index Terms—Graph regularization, hyperspectral unmixing, pointwise mutual information, sparse regression.

I. INTRODUCTION

WITH the increasing availability of spectral libraries, hyperspectral unmixing has been addressed under sparse regression techniques recently. Sparse unmixing (SU) aims at estimating fractional proportions (abundances) assuming that the spectral library, a collection of constituent spectra (endmembers), is known in advance [1]. In the linear SU framework, plenty of methods have been developed with various assumptions. The sparse unmixing via variable splitting and augmented Lagrangian (SUnSAL) [2], a leading work, has been proposed to solve a relaxed minimization problem using the alternating direction method of multipliers (ADMM) [3] strategy. Subsequently, many improvements to SUnSAL have been done. Collaborative SUnSAL (CLSUnSAL) [4] employs $l_{2,1}$ mixed norm to exploit the assumption that all pixels in an HSI share the same set of endmembers. The weighted mixed norm has been introduced in [5] to improve the performance of CLSUnSAL. The total variation (TV) regularizer, which takes into account spatial homogeneity, is included in SU formulation to integrate the spatial-contextual information of neighboring pixels [6]–[8]. To enhance the sparsity of abundances in both spatial and spectral domains, a double reweighted scheme is proposed in [7], [9]. The

spectral-spatial weighted SU (S²WSU) presents a framework for operating strong spatial correlation among features in HSI data [10]. Along with these works, superpixel segmentation based studies have also been developed to make use of spatial-contextual information [8], [11]–[13].

Although the TV regularizer considerably improves the performance of SU methods, it only deals with the similarity of first-order neighboring domains and disregards the relationships of other pixels in the image, even if they show similar features. It may also cause over-smoothed abundance maps in the transition areas. On the other hand, graph Laplacian (GL) regularizer, which is quite successful in representing the latent structure between pixels, has been investigated as an alternative to the TV in [12]–[16]. All these methods use the heat kernel [17] that only considers the spectral similarity of pixels for constructing similarity graphs. To take into account spatial-contextual information, the superpixel segmentation strategy has been employed in [12], [13] and hypergraph learning is used in [18], which explores the similarity of pixels in a local neighborhood. Additionally, the GL created by the heat kernel may not be reliable especially when the observed data is extremely contaminated by noise [16].

In this letter, we propose to construct a new similarity-weighting matrix for SU instead of the heat kernel to take full advantage of spatial-contextual information as well as spectral similarity in HSI. Inspired by their success in spectral clustering [19] and LiDAR fused unsupervised HSI segmentation [20], we build affinity matrices using pointwise mutual information (PMI), which operates on the principle that pixels belonging to the same object show higher statistical dependencies than pixels of distinct objects [19]. We incorporate PMI-based GL regularization and double reweighted ℓ_1 norm minimization to promote SU performance.

II. PROPOSED METHOD

In this section, we give the proposed method, but first, we provide the necessary background. Let $\mathbf{Y}$ represent the $L \times K$ matrix of pixels in an HSI, where L is the number of spectral bands and K is the number of mixed pixels. In the linear SU model,

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{L \times m}$ is the spectral library containing m material signatures, $\mathbf{X} \in \mathbb{R}^{m \times K}$ is the abundances and each column of this matrix, $\mathbf{x}_i$ ($i = 1, 2, \dots, K$) is the $m \times 1$ vector representing the abundances of the i^{th} pixel, and $\mathbf{N} \in \mathbb{R}^{L \times K}$ denotes a matrix of the noise. SUnSAL

Sefa Kucuk is with the Department of Electrical and Electronics Engineering, Erzurum Technical University, Erzurum, Turkey, 25050 (e-mail: sefa.kucuk@erzurum.edu.tr).

Seniha Esen Yuksel is with the Department of Electrical and Electronics Engineering, Hacettepe University, Ankara, Turkey, 06800 (e-mail: eyuksel@ee.hacettepe.edu.tr).

formulates SU as a convex relaxed optimization problem, replacing the ℓ_0 norm with the ℓ_1 norm in the original regularization, as follows:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X}\|_{1,1} \quad \text{s.t.} \quad \mathbf{X} \geq 0. \quad (2)$$

Here, $\|\cdot\|_F$ is the Frobenius norm, λ is the regularization parameter, and $\|\mathbf{X}\|_{1,1} = \sum_{i=1}^K \|\mathbf{x}_i\|_1$ with $\mathbf{x}_i$ being the i^{th} column of $\mathbf{X}$. SUNSAL solves (2) efficiently by ADMM. Table I lists the objective functions of the methods aforementioned in the Introduction, which will be used for comparative purposes in Section III. In Table I, $\text{TV}(\mathbf{X}) \equiv \sum_{\{i,j\} \in \kappa} \|\mathbf{x}_i - \mathbf{x}_j\|_1$ represents the TV regularizer, where κ shows the set of horizontal and vertical pixel neighbors. On the other hand, $\text{Tr}(\mathbf{XLX}^T)$ stands for the GL regularizer, where $\mathbf{L}$ is the Laplacian matrix. The $\mathbf{W}_1$ and $\mathbf{W}_4 \in \mathbb{R}^{m \times m}$ matrices are used to promote nonzero row vectors of $\mathbf{X}$. The $\mathbf{W}_1$ matrix is designed as follows:

$$\mathbf{W}_1^{t+1} = \text{diag} \left(\frac{1}{\|\mathbf{X}^t(1,:) \|_2 + \epsilon}, \dots, \frac{1}{\|\mathbf{X}^t(m,:) \|_2 + \epsilon} \right) \quad (3)$$

where $\epsilon > 0$ and $\mathbf{X}^t(m,:)$ denotes the m^{th} row of the estimated $\mathbf{X}$ in the t^{th} iteration. The $\mathbf{W}_4$ matrix is also obtained with (3), replacing the ℓ_1 norm with the ℓ_2 norm. The $\mathbf{W}_2 \in \mathbb{R}^{m \times K}$ is employed to encourage sparsity in nonzero rows and defined as

$$\mathbf{W}_2^{t+1} = \frac{1}{|\mathbf{X}^t| + \epsilon}. \quad (4)$$

Finally, $\mathbf{W}_3 \in \mathbb{R}^{m \times K}$ is created by means of a function that explicitly utilizes the spatial-contextual information through the neighborhood system, such as the neighboring coverage [10]. An iterative reweighting scheme is adopted to update all weight matrices. The last column of Table I shows whether the methods make use of spatial-contextual information or not.

A. Pointwise Mutual Information

Mutual information (MI) is a measure of the mutual dependence between two random variables. PMI, on the other hand, only measures the mutual dependence between two realizations of random variables. Isola *et al.* [19] have achieved excellent spectral clustering results by modeling the association between neighboring image features through a statistical framework based on PMI. To calculate the statistical association of the feature such as color, intensity of two pixels in an image, the affinity matrix is constructed based on probability density estimation via PMI as follows:

$$\text{PMI}_\rho(\mathbf{f}_i, \mathbf{f}_j) = \log \frac{P(\mathbf{f}_i, \mathbf{f}_j)^\rho}{P(\mathbf{f}_i)P(\mathbf{f}_j)} \quad (5)$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ denotes the neighboring feature vectors of two pixels in the image. The joint probability $P(\mathbf{f}_i, \mathbf{f}_j)$ in (5) is defined as

$$P(\mathbf{f}_i, \mathbf{f}_j) = \frac{1}{Z} \sum_{d=d_0}^{\infty} w(d) p(\mathbf{f}_i, \mathbf{f}_j; d) \quad (6)$$

in which, $p(\mathbf{f}_i, \mathbf{f}_j; d)$ is the joint probability of features $\mathbf{f}_i$ and $\mathbf{f}_j$ occurring at a Euclidean distance of d pixels apart, Z

TABLE I
A BRIEF DESCRIPTION OF THE SU METHODS

Method	Optimization Problem	Spatial Information
SUNSAL [2]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ \mathbf{X}\ _{1,1}$ s.t. $\mathbf{X} \geq 0$	–
SUNSAL-TV [6]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ \mathbf{X}\ _{1,1}$ + $\lambda_{\text{TV}} \text{TV}(\mathbf{X})$ s.t. $\mathbf{X} \geq 0$	✓
CLSunSAL [4]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ \mathbf{X}\ _{2,1}$ s.t. $\mathbf{X} \geq 0$	–
DRSU [9]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ (\mathbf{W}_1 \mathbf{W}_2) \odot \mathbf{X}\ _{1,1}$ s.t. $\mathbf{X} \geq 0$	–
DRSU-TV [7]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ (\mathbf{W}_1 \mathbf{W}_2) \odot \mathbf{X}\ _{1,1}$ + $\lambda_{\text{TV}} \text{TV}(\mathbf{X})$ s.t. $\mathbf{X} \geq 0$	✓
S ² WSU [10]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ (\mathbf{W}_1 \mathbf{W}_3) \odot \mathbf{X}\ _{1,1}$ s.t. $\mathbf{X} \geq 0$	✓
DRSGHU [16]	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{AX} - \mathbf{Y}\ _F^2 + \lambda \ (\mathbf{W}_4 \mathbf{W}_2) \odot \mathbf{X}\ _{1,1}$ + $\frac{\lambda_{\text{GL}}}{2} \text{Tr}(\mathbf{XLX}^T)$ s.t. $\mathbf{X} \geq 0$	–

is a normalization constant, and $w(\cdot)$ is a Gaussian weighting function that decreases monotonically with distance d . The multiple pairs are randomly sampled from all location in the image to compute (6). The probabilities $P(\mathbf{f}_i)$ and $P(\mathbf{f}_j)$ is obtained by taking the marginals of $P(\mathbf{f}_i, \mathbf{f}_j)$. The ρ is an adjustable parameter to optimize spectral clustering performance [19]. In this letter, we also treat it as a free parameter to get the best SU results.

B. Proposed Unmixing Model

In this letter, we build the similarity-weighting matrix via PMI by modeling the association between neighboring image features, and then use them in the GL regularizer for SU. The objective function of the proposed method is:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|(\mathbf{W}_1 \mathbf{W}_2) \odot \mathbf{X}\|_{1,1} + \frac{\lambda_{\text{GL}}}{2} \text{Tr}(\mathbf{XLX}^T) \quad \text{s.t.} \quad \mathbf{X} \geq 0 \quad (7)$$

where $\lambda \geq 0$ and $\lambda_{\text{GL}} \geq 0$ are regularization parameters, the operator $\odot$ is the component-wise product of two variables, and $\text{Tr}(\cdot)$ denotes the trace of a matrix. In our model, we also use the $\mathbf{W}_1$ and $\mathbf{W}_2$ weight matrices as in (3) and (4), respectively to promote sparsity of abundances. We add the last term to commit the fact that pixels with similar spectra in spatial proximity should have similar abundances. In the last term, $\mathbf{L} = \mathbf{D} - \mathbf{\Psi}$ denotes the GL matrix where $\mathbf{\Psi}$ is the affinity matrix and $\mathbf{D}$ is a diagonal matrix that can be calculated as $\mathbf{D}_{ii} = \sum_j \mathbf{\Psi}_{ij}$. We construct the affinity matrix as follows:

$$\mathbf{\Psi}_{i,j} = e^{\sum_{k=1}^2 \text{PMI}_\rho(\mathbf{f}_i^k, \mathbf{f}_j^k)} \quad (8)$$

in which $\mathbf{f}_i^1$ and $\mathbf{f}_i^2$ represent the spectral and spatial information of the i^{th} pixel, respectively. Since there is no correlation between the spectral properties and spatial positions of materials in a scene, we calculate PMIs for the spectral and spatial information separately. However, while creating the affinity matrix, we fuse PMIs as in (8).

We utilize the ADMM framework to solve (7). Specifically, we convert (7) into its equivalent form as

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4} & \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|(\mathbf{W}_1 \mathbf{W}_2) \odot \mathbf{V}_2\|_{1,1} \\ & + \frac{\lambda_{\text{GL}}}{2} \text{Tr}(\mathbf{V}_3 \mathbf{L} \mathbf{V}_3^T) + \iota_{R+}(\mathbf{V}_4) \\ \text{s.t. } & \mathbf{V}_1 = \mathbf{A} \mathbf{U}, \mathbf{V}_2 = \mathbf{U}, \mathbf{V}_3 = \mathbf{U}, \mathbf{V}_4 = \mathbf{U} \end{aligned} \quad (9)$$

where $\iota_{R+}(\mathbf{X}) = \sum_{i=1}^K \iota_{R+}(\mathbf{x}_i)$ is the indicator function ($\iota_{R+}(\mathbf{x}_i)$ is zero when $\mathbf{x}_i$ belongs to the nonnegative orthant and $+\infty$ otherwise). Then, we formulate the augmented Lagrangian as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4) &= \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 \\ &+ \lambda \|(\mathbf{W}_1 \mathbf{W}_2) \odot \mathbf{V}_2\|_{1,1} + \frac{\lambda_{\text{GL}}}{2} \text{Tr}(\mathbf{V}_3 \mathbf{L} \mathbf{V}_3^T) + \iota_{R+}(\mathbf{V}_4) \\ &+ \frac{\mu}{2} \|\mathbf{A} \mathbf{U} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2 + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V}_2 - \mathbf{D}_2\|_F^2 \\ &+ \frac{\mu}{2} \|\mathbf{U} - \mathbf{V}_3 - \mathbf{D}_3\|_F^2 + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V}_4 - \mathbf{D}_4\|_F^2 \end{aligned} \quad (10)$$

where μ is an augmented Lagrangian parameter and $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$, and $\mathbf{D}_4$ denote the Lagrange multipliers. The procedures for solving the optimization problem with the proposed method through ADMM are shown Algorithm 1. We perform an inner-outer looping scheme as used in [10] to make the convergence of the algorithm better. In the experimental section, we set the number of iterations of the inner and outer loops as $k = 10$ and $l = 50$, respectively.

III. EXPERIMENTAL ANALYSIS

In this section, we illustrate the performance of the proposed method on one synthetic and two real HSI data sets. To quantitatively evaluate the SU results, we use the signal-to-reconstruction error (SRE) which is defined as $\text{SRE}(\text{dB}) = 10 \log_{10} \left(\frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2} \right)$, the root mean square error - RMSE, the running times, and the *sparsity*, which measures the proportion of the number of elements in $\hat{\mathbf{X}}$ that are greater than 0.005 with regard to all elements [10]. $\hat{\mathbf{X}}$ denotes the estimated abundance by SU methods and $\mathbf{X}$ is the true abundance.

We can derive PMI($\mathbf{f}_i, \mathbf{f}_j$) by modeling $P(\mathbf{f}_i, \mathbf{f}_j)$ according to the internal image statistics. To do this, we use a nonparametric kernel density estimator, and place Epanechnikov kernels at randomly selected 10000 sample points per data set. After sampling the position of a randomly selected pixel in the image, we sample the features $\mathbf{f}_i$ and $\mathbf{f}_j$ from image locations around the selected pixel, such that $\mathbf{f}_i$ and $\mathbf{f}_j$ are d pixels apart. The majority of sample pairs are within 4 pixels of each other. For the weighting function $w(d)$ in (6), we use a Gaussian with $\sigma = 0.01$. On the other hand, since HSI images consist of hundreds of spectral channels, plenty of time will be needed to learn $P(\mathbf{f}_i, \mathbf{f}_j)$ for spectral information. To prevent this case, we perform dimension reduction using the principal component analysis (PCA).

A. Evaluation on the Synthetic Data Set

Here, we utilize the spectral library $\mathbf{A} \in \mathbb{R}^{224 \times 240}$, which contains 240 spectral signatures with 224 bands from

Algorithm 1 Pseudocode of the proposed method

Input: Hyperspectral data $\mathbf{Y} \in \mathbb{R}^{L \times K}$; Dictionary $\mathbf{A} \in \mathbb{R}^{L \times m}$; GL matrix $\mathbf{L} \in \mathbb{R}^{K \times K}$; $\lambda, \lambda_{\text{GL}}$ and μ

Initialization: $k = 0, l = 0, \mathbf{U}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, \mathbf{V}_3^{(0)}, \mathbf{V}_4^{(0)}, \mathbf{D}_1^{(0)}, \mathbf{D}_2^{(0)}, \mathbf{D}_3^{(0)}, \mathbf{D}_4^{(0)}$

Repeat:

$$\mathbf{W}_1^l \leftarrow \text{diag} \left[\frac{1}{\|(\mathbf{U}^{(l)} - \mathbf{D}_2^{(l)})_{(1,:)}\|_2 + \varepsilon}, \dots, \frac{1}{\|(\mathbf{U}^{(l)} - \mathbf{D}_2^{(l)})_{(m,:)}\|_2 + \varepsilon} \right]$$

$$\mathbf{W}_2^l \leftarrow \frac{1}{\|\mathbf{U}^{(l)} - \mathbf{D}_2^{(l)}\|_2 + \varepsilon}$$

Repeat:

$$\mathbf{U}^{(k+1)} \leftarrow (\mathbf{A}^T \mathbf{A} + 3\mathbf{I})^{-1} [\mathbf{A}^T (\mathbf{V}_1^{(k)} + \mathbf{D}_1^{(k)}) + (\mathbf{V}_2^{(k)} + \mathbf{D}_2^{(k)}) + (\mathbf{V}_3^{(k)} + \mathbf{D}_3^{(k)}) + (\mathbf{V}_4^{(k)} + \mathbf{D}_4^{(k)})]$$

$$\mathbf{V}_1^{(k+1)} \leftarrow \frac{1}{1+\mu} [\mathbf{Y} + \mu (\mathbf{A} \mathbf{U}^{(k+1)} - \mathbf{D}_1^{(k)})]$$

$$\mathbf{V}_2^{(k+1)} \leftarrow \text{soft} \left(\mathbf{U}^{(k+1)} - \mathbf{D}_2^{(k)}, \frac{\lambda}{\mu} \mathbf{W}_1^{(l)} \mathbf{W}_2^{(l)} \right)$$

$$\mathbf{V}_3^{(k+1)} \leftarrow \mu (\mathbf{U}^{(k+1)} - \mathbf{D}_3^{(k)}) (\lambda_{\text{GL}} \mathbf{L} + \mu \mathbf{I})^{-1}$$

$$\mathbf{V}_4^{(k+1)} \leftarrow \max \{ \mathbf{U}^{(k)} - \mathbf{D}_4^{(k)}, 0 \}$$

Update Lagrange multipliers:

$$\mathbf{D}_1^{(k+1)} \leftarrow \mathbf{D}_1^{(k)} - (\mathbf{A} \mathbf{U}^{(k+1)} - \mathbf{V}_1^{(k+1)})$$

$$\mathbf{D}_2^{(k+1)} \leftarrow \mathbf{D}_2^{(k)} - (\mathbf{U}^{(k+1)} - \mathbf{V}_2^{(k+1)})$$

$$\mathbf{D}_3^{(k+1)} \leftarrow \mathbf{D}_3^{(k)} - (\mathbf{U}^{(k+1)} - \mathbf{V}_3^{(k+1)})$$

$$\mathbf{D}_4^{(k+1)} \leftarrow \mathbf{D}_4^{(k)} - (\mathbf{U}^{(k+1)} - \mathbf{V}_4^{(k+1)})$$

Update iteration: $k \leftarrow k + 1$

$$\mathbf{U}^{(l+1)} \leftarrow \mathbf{U}^{(k+1)}$$

Update iteration: $l \leftarrow l + 1$

Until: some stopping criterion is satisfied.

the United States Geological Survey (USGS) library [21]. We create the synthetic data set (SDS) by randomly selecting 5 spectra from library $\mathbf{A}$ as the constituent endmembers and using the abundance maps of 75×75 pixels given in [6]. Then, we corrupt the SDS by different levels of Gaussian noise (SNR = 20, 30, 40 and 50 dB). For the SDS, the sum of the first 7 largest eigenvalues constitutes 99.90% of the total of all eigenvalues. Based on this result, we get the PMIs for spectral information using the first 7 largest principal components of the data set. Experimentally, we obtain the highest SRE when ρ is 2 on the validation set, so we set ρ to 2 for all experiments. We share Fig. 1 as an instance to illustrate the effectiveness of PMI in spectral clustering. As can be seen from Fig. 1, crisp boundaries of the scene have been successfully found. Table II compares the quantitative analyses of SU performances with increasing levels of SNR. The reported values have been achieved with the optimal parameter(s) given in parameter lines. As highlighted in Table II, the proposed method is superior to the compared algorithms except for the DRS GHU in all metrics. On the other hand, our method achieves much better results than the DRS GHU at very low SNR levels (20

Fig. 1. (a) Pseudo-coloring of the SDS with bands 20, 40, 110 at 30 dB. (b) Crisp boundaries of the scene.

TABLE II
PERFORMANCE COMPARISON OF SU METHODS ON THE SDS

	SUNSAL	SUNSAL-TV	CLSUNSAL	DRSU	DRSU-TV	S ² WSU	DRSGHU	Ours
SNR = 20dB	RMSE	0.0299	0.0087	0.0175	0.0202	0.0079	0.0110	0.0027
	Sparsity	0.0652	0.0546	0.0524	0.0297	0.0528	0.0351	0.0276
	SRE(dB)	1.2492	12.0022	5.9118	4.6810	12.8104	9.9273	22.2826
	Parameter(s)	$\lambda = 2e-2$	$\lambda = 5e-4$ $\lambda_{TV} = 5e-2$	$\lambda = 1e2$	$\lambda = 1e-1$	$\lambda = 1e-3$ $\lambda_{TV} = 1e-2$	$\lambda = 1e-2$	$\lambda = 1e0$ $\lambda_{GL} = 1e3$ $\lambda_{GL} = 3e-1$
SNR = 30dB	RMSE	0.0156	0.0029	0.0085	0.0035	0.0017	0.0029	0.0010
	Sparsity	0.0737	0.0335	0.0705	0.0240	0.0235	0.0208	0.0208
	SRE(dB)	6.9193	21.6538	12.1315	19.9771	26.3446	21.5871	30.6880
	Parameter(s)	$\lambda = 5e-3$	$\lambda = 5e-4$ $\lambda_{TV} = 1e-2$	$\lambda = 1e0$	$\lambda = 5e-3$	$\lambda = 2e-3$ $\lambda_{TV} = 2e-3$	$\lambda = 5e-3$	$\lambda = 3e-2$ $\lambda_{GL} = 1e3$ $\lambda_{GL} = 1e-1$
SNR = 40dB	RMSE	0.0053	0.0012	0.0030	0.0009	0.0006	0.0009	0.0003
	Sparsity	0.0585	0.0207	0.0340	0.0199	0.0200	0.0200	0.0199
	SRE(dB)	16.2319	29.4592	21.1182	31.6258	35.5923	31.8193	42.2793
	Parameter(s)	$\lambda = 1e-3$	$\lambda = 5e-4$ $\lambda_{TV} = 5e-3$	$\lambda = 3e-1$	$\lambda = 1e-3$	$\lambda = 6e-4$ $\lambda_{TV} = 4e-4$	$\lambda = 5e-4$	$\lambda = 1e-2$ $\lambda_{GL} = 1e3$ $\lambda_{GL} = 3e-2$
SNR = 50dB	RMSE	0.0017	0.0004	0.0010	0.0003	0.0002	0.0003	0.0001
	Sparsity	0.0289	0.0199	0.0202	0.0199	0.0199	0.0199	0.0199
	SRE(dB)	26.0822	37.9724	30.9029	41.7651	44.7463	41.7099	49.8842
	Parameter(s)	$\lambda = 7e-4$	$\lambda = 5e-4$ $\lambda_{TV} = 5e-4$	$\lambda = 5e-2$	$\lambda = 2e-4$	$\lambda = 1e-4$ $\lambda_{TV} = 1e-4$	$\lambda = 1e-4$	$\lambda = 1e-2$ $\lambda_{GL} = 1e3$ $\lambda_{GL} = 1e-2$

TABLE III
COMPARISON OF SU RESULTS ON THE RDS1

Algorithms	SUNSAL	SUNSAL-TV	CLSUNSAL	DRSU	DRSU-TV	S ² WSU	DRSGHU	Ours
RMSE	0.0144	0.0144	0.0195	0.0138	0.0102	0.0101	0.0102	0.0082
Sparsity	0.0135	0.0198	0.0129	0.0089	0.0076	0.0059	0.0054	0.0049
SRE(dB)	8.4844	8.5406	5.9027	8.8587	11.5283	11.5780	11.5474	13.4004
Parameter(s)	$\lambda = 2e-3$	$\lambda = 2e-3$ $\lambda_{TV} = 1e-3$	$\lambda = 5e-3$	$\lambda = 1e-3$	$\lambda = 1e-2$ $\lambda_{TV} = 1e-2$	$\lambda = 2e-2$	$\lambda = 1e0$ $\lambda_{GL} = 1e3$	$\lambda = 1e-2$ $\lambda_{GL} = 1e-3$

and 30 dB). These results consolidate the argument that the heat kernel cannot form a reliable GL matrix for low SNR levels, as mentioned before.

B. Evaluation on Real Data Sets

Here, we demonstrate the performance of our method on two real data sets.

1) *Real Data Set 1 (RDS1)*: A subregion of 100×100 of Jasper Ridge data set (available online at <https://rslab.ut.ac.ir/data>) captured by AVIRIS is used. It has 198 spectral bands after the removal of low SNR and water-vapor absorption bands. There are four targets of interest in the scene: *tree*, *water*, *soil*, and *road* [22]. We use a spectral library containing 502 spectra following the procedure in [22]. We report the SU results in Table III along with the optimal regularization parameters under comparison. We also qualitatively compare the abundance maps estimated by the algorithms in Fig. 2. It is apparent from Table III that the proposed method outperforms the compared algorithms in all metrics. From Fig. 2 we can see that all SU algorithms are able to estimate abundance maps similar to the reference ground truth. Moreover, we can note that the proposed method obtains the most similar ones, conserving fine structures and texture.

2) *Real Data Set 2 (RDS2)*: A subscene of 204×151 of the Nevada Cuprite data set (available online at https://aviris.jpl.nasa.gov/data/free_data.html) which contains 188 spectral bands after removing the water absorption and low SNR bands is used. We create a spectral library $\mathbf{A} \in \mathbb{R}^{188 \times 498}$ from the USGS, removing the corresponding bands. Here, we qualitatively analyze the fractional abundance maps since true abundances of the RDS2 are not available. We benefit from the classification maps produced by the Tricorder 3.3 software product [23] as a reference. Regularization parameters of the compared methods are given in Table IV. We could not run DRSGHU on the

TABLE IV
PARAMETERS FOR THE SU ALGORITHMS APPLIED ON THE RDS2

Algorithms	SUNSAL	SUNSAL-TV	CLSUNSAL	DRSU	DRSU-TV	S ² WSU	Ours
Parameter(s)	$\lambda = 1e-4$	$\lambda = 5e-3$ $\lambda_{TV} = 1e-3$	$\lambda = 1e-3$	$\lambda = 5e-4$	$\lambda = 5e-4$ $\lambda_{TV} = 1e-4$	$\lambda = 1e-4$	$\lambda = 1e-5$ $\lambda_{GL} = 1e-3$

RDS2 because we could not take the inverse of the high-dimensional dense matrix produced using the heat kernel.

We present the estimated abundance maps for *Chalcodony* in Fig. 3 achieved by SU methods. As shown by the abundance maps, the regions with high abundances estimated by all the algorithms are consistent with the classification map. Herein, both the proposed method and other methods have obtained reasonable unmixing results on the RDS2. We note from the figures that TV-based methods smooth the details of abundances. It can be seen from Fig. 3 that the proposed method yields the most similar abundance map by eliminating outliers without smoothing the details.

C. Running Time Analysis

The processing time of the considered methods on three different data sets are shown together in Table V. Run times indicate the processing times needed to estimate the abundances of a mixed pixel. Therefore, the elapsed times have been divided by the total number of pixels in the data sets. We have implemented all the algorithms using MATLAB R2017a on a personal computer equipped with an Intel Core i7-4700HQ CPU with 2.40 GHz and 16 GB of RAM. The processing time of the algorithms is proportional to the additional terms added to the SU formulation, as updating the additional matrices in optimization steps increases the running time. Therefore the TV regularized SUNSAL-TV and DRSU-TV, and the GL regularized DRSGHU and the proposed method require more processing time. The factor that determines the processing time for the GL-based method is the size of the GL matrix, that is, the total number of pixels in the data. As can be seen from the first line of Table V, our method requires less run time than two TV-based methods. However, the increase in the number of pixels in the data causes an increase in the processing time, as shown in the last line of Table V. Compared with the DRSGHU, our method requires less processing time due to

Fig. 2. Fractional abundance maps for the RDS1 estimated by SU methods. Here, gt stands for the ground truth of the relevant materials.

Fig. 3. Fractional abundance maps for *Chalcedony*, (a) The distribution produced by Tricorder software, (b) SUNSAL, (c) SUNSAL-TV, (d) CLSUnSAL, (e) DRSU, (f) DRSU-TV, (g) S^2WSU , and (h) OURS.

TABLE V

PROCESSING TIMES (PER PIXEL IN MILLISECONDS) OF THE METHODS

Algorithms	SUNSAL	SUNSAL-TV	CLSUnSAL	DRSU	DRSU-TV	S^2WSU	DRSGHU	Ours
SDS	0.410	31.503	11.828	4.036	32.697	6.277	54.015	20.56
RDS1	12.817	56.973	27.701	7.45	59.835	11.874	174.997	58.331
RDS2	4.214	70.236	26.196	7.104	71.015	10.526	—	164.432

the fact that affinity matrices generated with PMI have a block and sparse structure.

IV. CONCLUSION

In this letter, we have introduced a GL-based unmixing framework for HSI data analysis. We demonstrate that PMI affinity matrices could be added to the SU formulation as a GL regularizer to exploit the rich spectral content and efficient spatial information. We have confirmed that PMI, obtained by statistical modeling of data-driven information, is quite successful in characterizing the hidden structure in HSI. We have proved the effectiveness of the proposed method with experiments on different data sets and obtained qualitatively and quantitatively superior results compared to other competing algorithms.

REFERENCES

- [1] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.
- [2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *2010 2nd WHISPERS*, pp. 1–4.
- [3] 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., January 2011, vol. 3, no. 1.
- [4] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Remote Sens.*, vol. 52, no. 1, pp. 341–354, 2014.
- [5] C. Y. Zheng, H. Li, Q. Wang, and C. L. Philip Chen, "Reweighted sparse regression for hyperspectral unmixing," *IEEE Trans. Remote Sens.*, vol. 54, no. 1, pp. 479–488, 2016.
- [6] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, 2012.
- [7] R. Wang, H.-C. Li, A. Pizurica, J. Li, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using double reweighted sparse regression and total variation," *IEEE Geo. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1146–1150, 2017.
- [8] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Trans. Remote Sens.*, vol. 59, no. 1, pp. 629–647, 2021.
- [9] R. Wang, H.-C. Li, W. Liao, and A. Pizurica, "Double reweighted sparse regression for hyperspectral unmixing," in *IEEE IGARSS*, 2016, pp. 6986–6989.
- [10] S. Zhang, J. Li, H.-C. Li, C. Deng, and A. Plaza, "Spectral-spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Trans. Remote Sens.*, vol. 56, no. 6, pp. 3265–3276, 2018.
- [11] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, and C. Richard, "A fast multiscale spatial regularization for sparse hyperspectral unmixing," *IEEE Geo. Remote Sens. Lett.*, vol. 16, no. 4, pp. 598–602, 2019.
- [12] T. Ince, "Superpixel-based graph laplacian regularization for sparse hyperspectral unmixing," *IEEE Geo. Remote Sens. Lett.*, pp. 1–5, 2020.
- [13] —, "Double spatial graph laplacian regularization for sparse unmixing," *IEEE Geo. Remote Sens. Lett.*, pp. 1–5, 2021.
- [14] R. Ammanouil, A. Ferrari, and C. Richard, "A graph laplacian regularization for hyperspectral data unmixing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2015, pp. 1637–1641.
- [15] Y. Xu, F. Fang, and G. Zhang, "Similarity-guided and ℓ_p -regularized sparse unmixing of hyperspectral data," *IEEE Geo. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2311–2315, 2015.
- [16] S. Wang, T.-Z. Huang, X.-L. Zhao, G. Liu, and Y. Cheng, "Double reweighted sparse regression and graph regularization for hyperspectral unmixing," *Remote Sensing*, vol. 10, no. 7, p. 1046, 2018.
- [17] S. Xiang, F. Nie, and C. Zhang, "Semi-supervised classification via local spline regression," *IEEE Trans. on Pattern Anal. and Machine Intel.*, vol. 32, no. 11, pp. 2039–2053, 2010.
- [18] P. Jia, M. Zhang, and Y. Shen, "Hypergraph learning and reweighted ℓ_1 -norm minimization for hyperspectral unmixing," *IEEE JSTAR*, vol. 12, no. 6, pp. 1898–1904, 2019.
- [19] P. Isola, D. Zoran, D. Krishnan, and E. H. Adelson, "Crisp boundary detection using pointwise mutual information," in *European Conf. on Computer Vision*. Springer Publishing, 2014, pp. 799–814.
- [20] O. Torun and S. E. Yuksel, "Unsupervised segmentation of lidar fused hyperspectral imagery using pointwise mutual information," *Int. Jour. of Remote Sens.*, vol. 42, no. 17, pp. 6461–6476, 2021.
- [21] R. N. Clark, G. A. Swayze, R. A. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library splib06a," US Geological Survey, Tech. Rep., 2007.
- [22] L. Qi, J. Li, Y. Wang, Y. Huang, and X. Gao, "Spectral-spatial-weighted multiview collaborative sparse unmixing for hyperspectral images," *IEEE Trans. Remote Sens.*, vol. 58, no. 12, pp. 8766–8779, 2020.
- [23] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," vol. 108, no. E12, 2003.