HYPERSPECTRAL DATA UNMIXING WITH GRAPH-BASED REGULARIZATION

Rita Ammanouil, André Ferrari, Cédric Richard

Lab. J.-L. Lagrange, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur
{rita.ammanouil, andre.ferrari, cedric.richard}@unice.fr

ABSTRACT

This work proposes to solve the unmixing problem in a graph setting. The hyperspectral image is mapped to a weighted graph where every pixel spectrum is represented by a node, and *similar* nodes are connected by weighted edges. A graph-based Total Variation framework is incorporated within the unmixing problem. The graph topology allows to promote smoothness over similar data in arbitrary neighborhood structures, while the weights allow to preserve details in the reconstructed image. The resulting constrained optimization problem is convex and is solved using an Alternating Direction Method of Multipliers (ADMM). The particular variable splitting allows to obtain closed form solutions for each ADMM sub-problem and avoids encountering more complicated Sylvester equations. Finally, experiments are conducted using synthetic and real hyperspectral data to demonstrate the effectiveness of the proposed algorithm.

***Index Terms*—** Hyperspectral images, sparse unmixing, non-local Total Variation, graph regularization.

1. INTRODUCTION

Hyperspectral sensors provide hundreds of spectrally contiguous measurements of a scene over the visible, near infrared and short wave infrared bands [1]. Each pixel is described by a thorough sampled spectrum allowing to identify its composition. Hyperspectral images usually have low spatial resolution, in the sense that a pixel covers a large spatial area. It is very probable that the area covered by the pixel contains more than one material. Linear unmixing [2] assumes that the observed pixel spectrum is a convex combination of some constituent material spectra. The constituent material spectra are known as the endmembers, and their proportions as the abundances.

We propose to generate a weighted graph from the hyperspectral data that encodes spatial and spectral similarities between the pixels. Each pixel in the hyperspectral image is represented by a node, and edges are used to connect spatially and spectrally similar nodes. We then perform sparse unmixing [3] while taking into account the additional relational information provided by the graph topology. More precisely, if two pixels are connected by an edge then it is assumed that their reconstructed spectra should be similar. Using tools of discrete calculus on graphs [4], we penalize the discrepancies between the estimated spectra of connected pixels via a graph-based Total Variation. The idea of Graph-based regularization has been widely used in the literature. Sometimes it is referred to as the Laplacian regularization when it is performed with the ℓ_2-norm. It has been successfully applied in many fields such as non-local image denoising [5–7], and semi-supervised learning [8–11].

This work was partly supported by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12-BS03-003), and the regional council of Provence-Alpes-Côte d'Azur.

The proposed strategy is closely related to the work in [12] where the authors use a Total Variation (TV) regularization for sparse ℓ_1-norm regularized unmixing. Similarly, this communication advocates the use of the graph TV regularization to perform $\ell_{2,1}$-norm regularized unmixing [3, 13]. However, TV is restricted to the assumption of local spatial similarity, and only relates a pixel to its four neighbors. In addition to this, TV penalizes all high variations between connected spectra despite their natural presence in real images (for example contours). We use an arbitrary graph which makes it possible to use the spatial and spectral information to make better decisions about the links between pixels and thus improve the prior at each pixel. Very recently, the authors of [14] and [15] used the graph Laplacian regularization for sparse $\ell_{1/2}$-norm regularized Non negative Matrix Factorization (NMF) within the context of blind unmixing. Their algorithm performs alternate minimization in order to simultaneously estimate the endmembers and the abundances. In this work, we use the ADMM algorithm [16] which allows to take into account the sum-to-one and positivity constraints for the abundances, and a Group lasso regularizer frequently incorporated in unmixing to allow the use of large libraries of endmembers. Finally, unlike the two previous approaches, Total Variation in this paper is imposed on the reconstructed spectra rather than on the abundances. This choice, which is more intuitive in our case since the graph is constructed using the image itself, is validated in the experiments. The resulting optimization problem can be solved efficiently using the ADMM. For a review of methods that incorporate other spatial or spectral-spatial information in the unmixing problem, the reader is referred to [17].

The paper is organized as follows. Section 2 introduces the graph Total Variation regularization within the unmixing problem. Section 3 is devoted to the ADMM solution. Finally, section 4 presents experiments using synthetic and real hyperspectral images.

2. GRAPH-BASED UNMIXING

Let us first introduce the linear mixing model which can be expressed in matrix form as

$$\boldsymbol{S} = \boldsymbol{R}\boldsymbol{A} + \boldsymbol{E} \tag{1}$$

with $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N)$, $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M)$, $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M)^\top$. Here, $\boldsymbol{s}_j$ is the L-dimensional spectrum of the j-th pixel, L is the number of frequency bands, $\boldsymbol{r}_i$ is L-dimensional spectrum of the i-th endmember, M denotes the number of endmembers, $\boldsymbol{a}_i$ is the M-dimensional abundance map of the i-th endmember, N is the number of pixels in the image, and $\boldsymbol{E}$ is an additive Gaussian noise. All vectors are column vectors. Model (1) means that the (i, j)-th entry A_{ij} of $\boldsymbol{A}$ is the abundance of endmember $\boldsymbol{r}_i$ in pixel $\boldsymbol{s}_j$. As mentioned previously, two constraints are usually imposed on the abundances, the non-negativity and sum-to-one constraints: $A_{ij} \geq 0$ for all (i, j), and $\sum_{i=1}^{M} A_{ij} = 1$ for all j, respectively. When $\boldsymbol{R}$ is a

large dictionary of endmembers, it is also required that $\boldsymbol{A}$ only has a few rows different from zero, those corresponding to the actual endmembers present in $\boldsymbol{S}$. To summarize, the unmixing problem requires that $\boldsymbol{RA}$ matches $\boldsymbol{S}$, and that $\boldsymbol{A}$ only has a few rows different from zero, in addition to the non-negativity and sum-to-one constraints. This leads to the following optimization problem,

$$\begin{array}{ll} \min_{\boldsymbol{A}} & \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{RA}\|_{\mathrm{F}}^2+\mu\sum_{k=1}^{M}\|\boldsymbol{a}_k\|_2 \\ \text{subject to} & A_{ij}\geq 0 \quad \forall\, i,j \\ & \sum_{i=1}^{M}A_{ij}=1 \quad \forall\, j \end{array} \tag{2}$$

where μ is a tuning parameter controlling the strength of the sparse group-lasso regularization. Before introducing the graph-based prior into problem (2), we first give some notations and describe the strategy used for generating a meaningful graph from the hyperspectral image. Let $\mathcal{G}=(\mathcal{V},\mathcal{E},\boldsymbol{W})$ be the image graph where $\mathcal{V}=\{v_1,\cdots,v_N\}$ is the set of vertices, $\boldsymbol{W}\in\mathbb{R}^{N\times N}$ is the affinity matrix, and $\mathcal{E}=\{e_{ij}|i\sim j\}$ is the set of (undirected) edges where $i\sim j$ means that vertices v_i and v_j are connected. Every vertex in $\mathcal{V}$ is associated with a pixel in the image, and the non negative entries W_{ij} of $\boldsymbol{W}$ are a measure of the similarity between the spectra corresponding to v_i and v_j. If two vertices v_i and v_j are similar, i.e. $W_{ij}\neq 0$, then they are connected by an edge denoted denoted by e_{ij}. There exists different techniques for generating a meaningful set of weighted edges to connect the nodes (see for example [18], chapter 4). Two straightforward options for defining the topology are either building a four neighborhood regular graph, as in classical TV, or a fully connected graph. The former seems like a natural possibility, however, it is not adapted to the underlying image. The latter is the most thorough choice, it takes advantage of all pairwise relationships. However, the number of edges can become prohibitive for large N. As a compromise, we connect each pixel to its first order spatial neighbors in addition to its k nearest neighbors among the other pixels in the image w.r.t. the spectral distance. Afterwards, a monotonically decreasing function of the spectral distance is used to assign a weight to each edge. Given $\mathcal{G}$, the gradient at v_i and band ℓ is defined as

$$\nabla_{v_i}^{\ell}\boldsymbol{S}=(W_{ij}|S_{li}-S_{lj}|)_j\in\mathbb{R}^N. \tag{3}$$

The TV of the hyperspectral image w.r.t. the graph is then given by the sum over all bands and all nodes of the gradient norms

$$J_{\mathcal{G}}(\boldsymbol{S})=\sum_{v_i\in\mathcal{V}}\sum_{\ell=1}^{L}\|\nabla_{v_i}^{\ell}\boldsymbol{S}\|_1=\sum_{e_{ij}\in\mathcal{E}}W_{ij}\|\boldsymbol{s}_i-\boldsymbol{s}_j\|_1. \tag{4}$$

This regularizer was used by [5] and [18] under different forms. The former defines it for univariate data and considers the ℓ_2 norm of the gradient. The latter considers any ℓ_p-norm and refers to it as the Basic Energy Model. In both references, it is used to perform image filtering by solving:

$$\min_{\widehat{\boldsymbol{S}}}\quad \frac{1}{2}\|\boldsymbol{S}-\widehat{\boldsymbol{S}}\|_{\mathrm{F}}^2+\lambda J_{\mathcal{G}}(\widehat{\boldsymbol{S}}), \tag{5}$$

where λ controls the strength of the graph TV regularization. Problem (5) aims at finding a smooth version $\widehat{\boldsymbol{S}}$ of $\boldsymbol{S}$. When $\mathcal{G}$ is a weighted non-local graph, the advantage over regular TV is twofold. The weights, which are proportional to the similarity between the two corresponding nodes, allow to smooth more over similar data and smooth less over dissimilar data. When two nodes are connected and similar, W_{ij} takes large values, preventing high variations between their estimates. Conversely, when W_{ij} is equal to zero or small, high variations are allowed. Instead of only using the local neighborhood of a pixel, spectrally similar pixels provided by a k nearest neighbor algorithm can also help improve the estimation [6]. This promotes collaborative unmixing between pixels located in similar regions across the image. Combining (2) and (5) yields

$$\begin{array}{ll} \min_{\boldsymbol{A}} & \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{RA}\|_{\mathrm{F}}^2+\mu\sum_{k=1}^{M}\|\boldsymbol{a}_k\|_2+\lambda J_{\mathcal{G}}(\boldsymbol{RA}) \\ \text{subject to} & A_{ij}\geq 0 \quad \forall\, i,j \\ & \sum_{i=1}^{M}A_{ij}=1 \quad \forall\, j. \end{array} \tag{6}$$

Unlike classical TV, the spatial regularization is imposed on the reconstructed spectra rather than directly on the abundances. The graph is intended to capture the features of the image. Thus, it seems more appropriate to impose this graph on the reconstructed image at each spectral band rather than on the abundance maps. In what follows, we use the matrix form to express (4):

$$J_{\mathcal{G}}(\boldsymbol{RA})=\|\boldsymbol{\Gamma}(\boldsymbol{RA})^{\top}\|_1 \tag{7}$$

where $\boldsymbol{\Gamma}^{\top}\in\mathbb{R}^{N\times|\mathcal{E}|}$ is graph incidence matrix [18]. Each row of $\boldsymbol{\Gamma}$ is indexed by an edge and has two nonzero elements, namely $\boldsymbol{\Gamma}_{e_{ij},i}=W_{ij}$ and $\boldsymbol{\Gamma}_{e_{ij},j}=-W_{ij}$, encoding which vertices are incident at the edge and its corresponding weight.

3. ADMM ALGORITHM

It is possible to efficiently solve problem (6) with the ADMM [16]. The solution is found by solving small sub-problems which in our case reduce to solving a set of linear equations or estimating proximity operators. We adopt the following variable splitting scheme:

$$\begin{array}{ll} \min\limits_{\boldsymbol{X},\boldsymbol{Y}_{1\to 3},\boldsymbol{Z}} & \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{RX}\|_{\mathrm{F}}^2+\mu\sum_{k=1}^{M}\|\boldsymbol{z}_k\|_2+\mathcal{I}(\boldsymbol{Z})+\lambda\|\boldsymbol{Y}_3\|_1 \\ \text{subject to} & \boldsymbol{BX}+\boldsymbol{CZ}=\boldsymbol{F} \\ & \boldsymbol{Y}_1=\boldsymbol{X}^{\top} \\ & \boldsymbol{Y}_2=\boldsymbol{Y}_1\boldsymbol{R}^{\top} \\ & \boldsymbol{Y}_3=\boldsymbol{\Gamma Y}_2 \end{array} \tag{8}$$

where

$$\boldsymbol{B}=\begin{pmatrix}\boldsymbol{I}\\ \mathbf{1}^{\top}\end{pmatrix},\ \boldsymbol{C}=\begin{pmatrix}-\boldsymbol{I}\\ \mathbf{0}^{\top}\end{pmatrix},\ \boldsymbol{F}=\begin{pmatrix}\mathbf{0}\\ \mathbf{1}^{\top}\end{pmatrix}.$$

$\boldsymbol{X},\boldsymbol{Y}_1,\boldsymbol{Y}_2,\boldsymbol{Y}_3,\boldsymbol{Z}$ are the ADMM variables. The constraints are imposed to ensure that problem (8) is equivalent to problem (6). The augmented Lagrangian for problem (8) is given by

$$\begin{aligned}\mathcal{L}_{\rho}(\boldsymbol{X},\boldsymbol{Y}_{1\to 3},\boldsymbol{Z},\boldsymbol{\Lambda},\boldsymbol{V}_{1\to 3})&=\frac{1}{2}\|\boldsymbol{S}-\boldsymbol{RX}\|_{\mathrm{F}}^2+\mu\sum_{k=1}^{M}\|\boldsymbol{z}_k\|_2\\ &+\mathcal{I}(\boldsymbol{Z})+\lambda\|\boldsymbol{Y}_3\|_1+\mathrm{tr}(\boldsymbol{\Lambda}^{\top}(\boldsymbol{BX}+\boldsymbol{CZ}-\boldsymbol{F}))+\mathrm{tr}(\boldsymbol{V}_1^{\top}(\boldsymbol{Y}_1-\boldsymbol{X}^{\top}))\\ &+\mathrm{tr}(\boldsymbol{V}_2^{\top}(\boldsymbol{Y}_2-\boldsymbol{Y}_1\boldsymbol{R}^{\top}))+\mathrm{tr}(\boldsymbol{V}_3(\boldsymbol{Y}_3-\boldsymbol{\Gamma Y}_2))+\frac{\rho}{2}\|\boldsymbol{BX}+\boldsymbol{CZ}-\boldsymbol{F}\|_F^2\\ &+\frac{\rho}{2}\|\boldsymbol{Y}_1-\boldsymbol{X}^{\top}\|_F^2+\frac{\rho}{2}\|\boldsymbol{Y}_2-\boldsymbol{Y}_1\boldsymbol{R}^{\top}\|+\frac{\rho}{2}\|\boldsymbol{Y}_3-\boldsymbol{\Gamma Y}_2\|_F^2\end{aligned}$$

where $\boldsymbol{\Lambda},\boldsymbol{V}_1,\boldsymbol{V}_2,\boldsymbol{V}_3$ are the Lagrange multipliers and ρ is the penalty parameter. At each ADMM iteration, the augmented Lagrangian is minimized w.r.t. each variable and the Lagrange multipliers are updated until a stopping criterion is satisfied.

$\boldsymbol{X}$ minimization step: After discarding the terms independent of $\boldsymbol{X}$ in the augmented Lagrangian, minimizing the augmented Lagrangian w.r.t. $\boldsymbol{X}$ reduces to a Least squares problem. The solution is obtained by solving a set of linear equations:

$$\begin{aligned}&(\boldsymbol{R}^{\top}\boldsymbol{R}+\rho\boldsymbol{B}^{\top}\boldsymbol{B}+\rho\boldsymbol{I})\boldsymbol{X}=\\ &\qquad\boldsymbol{R}^{\top}\boldsymbol{S}-\boldsymbol{B}^{\top}\boldsymbol{\Lambda}+\boldsymbol{V}_1^{\top}-\rho\boldsymbol{B}^{\top}(\boldsymbol{CZ}-\boldsymbol{F})+\rho\boldsymbol{Y}_1^{\top}.\end{aligned}$$

$\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ minimization step: Similarly to the first step, the $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ minimizations amount to solving sets of linear equations:

$$\boldsymbol{Y}_1(\boldsymbol{I}+\boldsymbol{R}^\top\boldsymbol{R}) = -\frac{1}{\rho}\boldsymbol{V}_1 + \frac{1}{\rho}\boldsymbol{V}_2\boldsymbol{R} + \boldsymbol{X}^\top + \boldsymbol{Y}_2\boldsymbol{R}, \quad (9)$$

$$(\boldsymbol{I}+\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma})\boldsymbol{Y}_2 = -\frac{1}{\rho}\boldsymbol{V}_2 + \frac{1}{\rho}\boldsymbol{\Gamma}^\top\boldsymbol{V}_3 + \boldsymbol{Y}_1\boldsymbol{R}^\top + \boldsymbol{\Gamma}^\top\boldsymbol{Y}_3. \quad (10)$$

Note that (9) and (10) are seperable row and colomn-wise respectively. Another variable splitting choice could lead to a more complex Sylvester equation non-seperable w.r.t. the unknown variables.

$\boldsymbol{Y}_3$ minimization step: Minimizing the augmented Lagrangian w.r.t. $\boldsymbol{Y}_3$ reduces to the well known lasso problem. The solution is obtained using soft thresholding:

$$\boldsymbol{Y}_3 = \text{soft}_{\frac{\lambda}{\rho}}(\boldsymbol{\Gamma}\boldsymbol{Y}_2 - \frac{1}{\rho}\boldsymbol{V}_3) \quad (11)$$

where $\text{soft}_\alpha(\cdot) = \text{sign}(\cdot)(|\cdot| - \alpha)_+$ is applied element wise and $(\cdot)_+ = \max(\boldsymbol{0}, \cdot)$.

$\boldsymbol{Z}$ minimization step: Minimizing the augmented Lagrangian w.r.t. $\boldsymbol{Z}$ reduces to solving the positively constrained group lasso problem:

$$\begin{cases} \boldsymbol{z}^* = \boldsymbol{0} & \text{if } \|(\boldsymbol{v})_+\|_2 < \alpha \\ \boldsymbol{z}^* = \left(1 - \frac{\alpha}{\|(\boldsymbol{v})_+\|_2}\right)(\boldsymbol{v})_+ & \text{otherwise} \end{cases} \quad (12)$$

where $\boldsymbol{v} = \boldsymbol{x} + \rho^{-1}\boldsymbol{\lambda}$, $\alpha = \rho^{-1}\mu$, $\boldsymbol{\lambda}$, $\boldsymbol{x}$ and $\boldsymbol{z}$ correspond to a row in $\boldsymbol{\Lambda}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ respectively.

Update the Lagrange multipliers: The last step at each ADMM iteration consists of updating the Lagrange multipliers

$$\begin{array}{ll} \boldsymbol{\Lambda}^{k+1} & = \boldsymbol{\Lambda}^k + \rho(\boldsymbol{B}\boldsymbol{X} + \boldsymbol{C}\boldsymbol{Z} - \boldsymbol{F}) \\ \boldsymbol{V}_1^{k+1} & = \boldsymbol{V}_1^k + \rho(\boldsymbol{Y}_1 - \boldsymbol{X}^\top) \\ \boldsymbol{V}_2^{k+1} & = \boldsymbol{V}_2^k + \rho(\boldsymbol{Y}_2 - \boldsymbol{Y}_1\boldsymbol{R}^\top) \\ \boldsymbol{V}_3^{k+1} & = \boldsymbol{V}_3^k + \rho(\boldsymbol{Y}_3 - \boldsymbol{\Gamma}\boldsymbol{Y}_2). \end{array} \quad (13)$$

Note that at each step, the most recent estimate of the variables and Lagrange multipliers is used.

4. EXPERIMENTS

The performance of the proposed approach was first evaluated using the synthetic data set described in [12]. It consist of a $75 \times 75 \times 224$ image with 25 homogeneous squares disposed in a grid fashion and lying on a homogeneous background. The data set was generated using 5 endmembers extracted from the USGS library which will be used in the simulations. As explained in section 2, we define the edge set of the graph by connecting a pixel to its four neighbors and to its 10 nearest neighbors where the spectral distance is measured with the ℓ_2-norm. We then use binary weights according to (14):

$$\begin{cases} W_{ij} = 1 & \text{if } \|\boldsymbol{s}_i - \boldsymbol{s}_j\|_2^2 < d_{\min}^2 \\ W_{ij} = 0 & \text{otherwise,} \end{cases} \quad (14)$$

where $d_{\min}^2$ represents the maximum squared spectral distance allowed between connected pixels. The performance of the proposed approach denoted by Graph TV is compared with FCLS, SUnSAL TV [12] and a TV regularized Collaborative unmixing [3] obtained by setting $\boldsymbol{Y}_2 = \boldsymbol{Y}_1$ in the third constraint of (8) and $d_{\min}^2 = \infty$ in (14). The penalty parameter was set to 0.05, and the maximum number of iterations to 200. Table 1 reports the best performances in terms of the Root Mean Square Error (RMSE $= (\frac{1}{NL} \times \|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_F^2)^{\frac{1}{2}}$) with the corresponding optimal pairs of regularization parameters. Graph TV requires tuning an additional parameter d_{min}^2 which is also reported in the table. All TV approaches outperformed FCLS. Graph TV had the lowest RMSE for all cases. The first row of Figure 1 shows the true abundance map of endmember $\boldsymbol{e}_1$, and the estimated maps obtained with Collaborative TV, SUnSAL TV and Graph TV with SNR = 30dB. It can be seen from these maps that both TV approaches and the proposed Graph TV estimated smooth abundance maps. However, the proposed approach was able to better recover the abundances of the squares in the second column. This is possibly due to the fact that the pixels in a square are connected to each other and disconnected from the background. This has prevented smoothing over the squares and making them disappear. Figure 2 shows the RMSE as a function of λ for the optimal μ in the case of the synthetic data set for a SNR of 30 dB.

Table 1. RMSE obtained with different values of SNR, with the optimal couple (μ; λ), ρ being set to 0.05.

	SNR 20 dB	SNR 30 dB	SNR 40 dB
FCLS	0.0262	0.0173	0.0101
SUnSAL TV	0.0156 (0.05; 0.05)	0.0075 ($5\ 10^{-3}$; 0.01)	0.0034 (10^{-3}; $5\ 10^{-3}$)
Collaborative TV	0.0151 (0.5; 0.05)	0.0071 (0.1; 0.01)	0.0028 (0.1; $5\ 10^{-3}$)
Graph TV	0.0101 (0.3; 0.01) $d_{min}^2 = 2.5$	0.0028 (0.1; 0.005) $d_{min}^2 = 0.3$	0.0010 (0.05; 10^{-3}) $d_{min}^2 = 0.05$

We also tested the proposed approach using real hyperspectral data, namely the Cuprite scene provided by NASA AVIRIS imaging spectrometer. The scene was captured over the mining district of Nevada with a spatial resolution of 17 meters over the wavelength interval $400 - 2500$ nm. After removing the water absorption bands, 188 bands were left for the analysis. We used the same USGS Library which contains the spectra of pure materials present in this scene. The second row of Figure 1 shows the abundances estimated using SUnSAL TV and the proposed approach for two endmembers. The tuning parameters λ and μ were both set to 10^{-3} for SUnSAL TV [12] and to 5×10^{-3} for the proposed algorithm, d_{min}^2 was set to 2.5. From the two endmember abundance maps, it can be seen that TV provided smoother results. However, the proposed approach was able to preserve relatively more details.

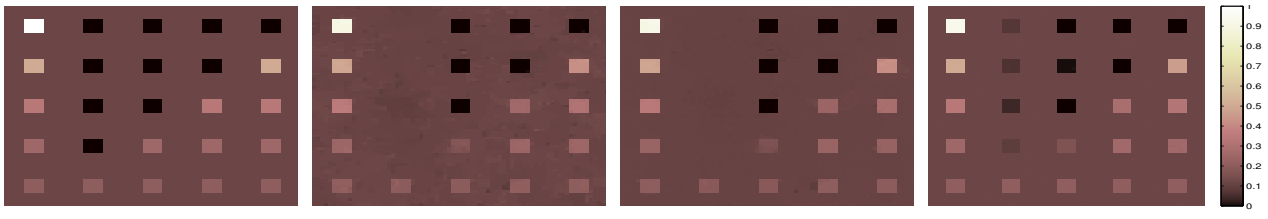
5. CONCLUSION

In this work, we incorporated a Graph based Total Variation within the unmixing problem. The proposed regularization provided smooth abundance maps while preserving details. Future work includes studying the potential of spatial-spectral weights for further improving the proposed approach.

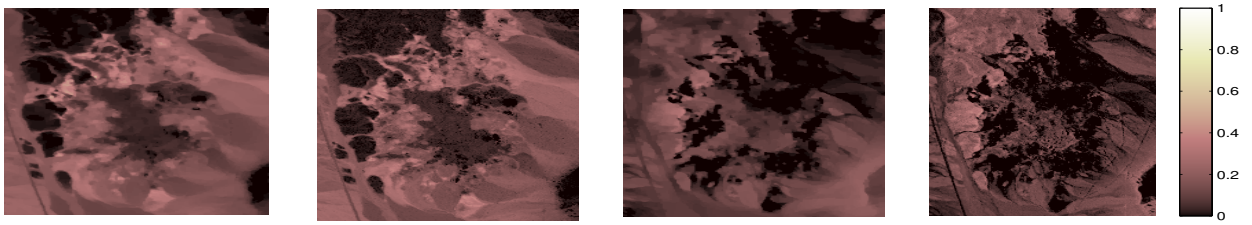
6. REFERENCES

[1] G. Shaw and H. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.



(a) True (b) Collaborative TV (c) SUnSAL TV (d) Graph TV



(e) SUnSAL TV (f) Graph TV (g) SUnSAL TV (h) Graph TV

Fig. 1. First row: Abundance maps for endmember 1 in the synthetic data set obtained with SNR $=$ 30 dB. The optimal parameters are reported in Table 1. Second row: Abundance maps for two endmembers in Cuprite obtained using SUnSAL TV and Graph TV.

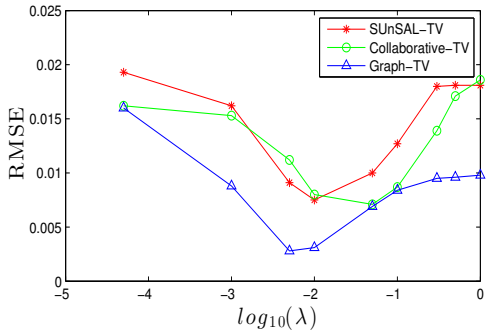


Fig. 2. RMSE between the true abundances and the estimated abundances for different values of the spatial regularisation.

[3] D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, February 2013.

[4] 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," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.

[5] G. Peyr, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Computer Vision–ECCV*, pp. 57–68. Springer, 2008.

[6] C. Couprie, L. Grady, L. Najman, J. C. Pesquet, and H. Talbot, "Dual constrained tv-based regularization on graphs," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1246–1273, 2013.

[7] F. Zhang and E. R. Hancock, "Graph spectral image smoothing using the heat kernel," *Pattern Recognition*, vol. 41, no. 11, pp. 3328–3342, 2008.

[8] G. Camps-Valls, T. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.

[9] A. Argyriou, M. Herbster, and M. Pontil, "Combining graph laplacians for semi–supervised learning," *NIPS*, 2005.

[10] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *Learning theory*, pp. 624–638. Springer, 2004.

[11] T. Zhang, A. Popescul, and B. Dom, "Linear prediction models with graph regularization for web-page categorization," in *Proc. ACM SIGKDD*, 2006, pp. 821–826.

[12] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[13] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "Blind and fully constrained unmixing of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5510–5518, 2014.

[14] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse nmf for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[15] L. Tong, J. Zhou, X. Bai, and Y. Gao, "Dual graph regularized nmf for hyperspectral unmixing," in *Proc. DICTA*, Nov. 2014.

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

[17] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sensing of Environment*, vol. 149, pp. 70–87, 2014.

[18] L. Grady, J. Leo, and J. Polimeni, *Discrete calculus: Applied analysis on graphs for computational science*, Springer, 2010.