ROBUST LINEAR UNMIXING WITH ENHANCED SPARSITY

A.Tiard[1], *L.Condat*[2], *L.Drumetz*[1,2], *J.Chanussot*[1,2], *W.Yin*[1] and *X.Zhu*[3]

[1]University of California at Los Angeles, UCLA, USA
[2]Univ. Grenoble Alpes, CNRS, GIPSA-lab, F-38000 Grenoble, France
[3]German Aerospace Center, and Tech. Univ. Munich, Germany
tiard@cs.ucla.edu, first.last@grenoble-inp.fr, wotaoyin@math.ucla.edu, xiao.zhu@dlr.de

ABSTRACT

Spectral unmixing is a central problem in hyperspectral imagery. It is usually assuming a linear mixture model. Solving this inverse problem, however, can be seriously impacted by a wrong estimation of the number of endmembers, a bad estimation of the endmembers themselves, the spectral variability of the endmembers or the presence of nonlinearities. These problems can result in a too large number of retained endmembers. We propose to tackle this problem by introducing a new formulation for robust linear unmixing enhancing sparsity. With a single tuning parameter the optimization leads to a range of behaviors: from the standard linear model (low sparsity) to a hard classification (maximal sparsity : only one endmember is retained per pixel). We solve the proposed new functional using a computationally efficient proximal primal dual method. The experimental study, including both realistic simulated data and real data demonstrates the versatility of the proposed approach.

Index terms - Hyperspectral Imagery, Linear Unmixing, Proximal Operators, Sparsity

1 INTRODUCTION

Hyperspectral imagery offers the possibility of pixel-level analysis, by splitting the electromagnetic spectrum in hundreds of narrow and contiguous spectral bands, typically covering the visible, near-infrared and short-wave infrared spectral domain [16]. The rich spectral content of the data allows to infer physical properties of the materials present in the scene. However, pixels are rarely - if at all - made up of only one material. On the contrary, a given pixel very often contains several different materials and only the mixture of these materials can be observed. Solving the inverse problem, i.e. estimating the pure materials present in the image(the *endmembers*) and their respective proportions in each pixel (the *abundances*) is referred to as the unmixing problem [3]. In a first step, one must find the number of endmembers to use [2], as well as their spectral signature [14], [12]. In a second step, one has to recover their abundances, for all pixels in the image. This is usually performed by assuming a linear mixture model (LMM), representing the observed spectrum as a linear combination of the endmembers. This is an ill-posed problem that will be solved by a constrained optimization.

Shortcomings of this traditionnal approach include the following issues: the sprectral variability of the endmembers [5], [9], the noise corrupting the data, the presence of potential nonlinearities ([8], [1], [10], [11]) may result in a wrong estimation of the number of endmembers and of their spectral signatures. Typically, we may end up with an *overestimation* of the number of endmembers. As a matter of fact, allowing for more endmembers adds flexibility to the model and improves the reconstruction of the observation. However, the pure materials may not be physically present in the scene [10], and the extracted ones may correspond to artifacts rather than physically interpretable components.

In this paper we propose to tackle this problem by enforcing an enhanced sparsity on the abundance maps while preserving a good reconstruction. Previous work aimed at adding robustness to the linear model [5] [6]. In this work, each pixel of the data is linearly explained, but the set of endmembers used is location-dependent. In other works, the endmembers are allowed to vary according to some constraints [9], [15], [17] and [7].

Our method also aims at providing robustness to the LMM, in light of the difficulty of estimating the pure materials in the scene. We propose a proof of concept for a new functional for the inversion of linear unmixing model, which includes enhanced sparsity scheme to reduce the impact of the poorly estimated

This work was supported by the ANR ASTRID program (Grant no. ANR-16-ASTR-0027-01 - APHYPIS project), CNRS (Grant no. PICS-USA 263484) and ERC (Grant no. ERC-2016-StG-714087 - So2Sat project)

endmembers. The new functional is minimized using a fast proximal splitting method.

The remaining sections of this paper are organized as follows. Section 2 introduces the formulation of the robust unmixing problem and associated functional. It also provides details on the implementation and the primal-dual algorithm. Results on both synthetic and real world datasets can be found in Section 3, with a critical discussion. Conclusions are drawn in section 4.

2 PROPOSED METHOD

2.1 Problem Formulation

A hyperspectral image Y with S bands and N pixels in vectorized form, is represented by a matrix of size $S \times N$. Let D be a matrix of size $S \times M$ whose columns are the spectral signatures of the endmembers.
Linear unmixing consists in finding the matrix A of size $M \times N$ of abundances: its element $a_{m,n}$ is the proportion of endmember number m assigned to pixel number n.

We propose the following formulation for Robust Linear Unmixing (RLU):

$$\begin{aligned} &minimize_A (1-\alpha) \parallel DA - Y \parallel^2 + \alpha \langle A,C \rangle + \mathcal{R}(A) \\ &s.t. \quad a_{\bullet,n} \in \Delta_M, \forall n = 1,\dots,N \end{aligned} \tag{1}$$

where Δ_M is the simplex over $\mathbb{R}^M$, $a_{\bullet,n}$ is the n-th column of A. $\mathcal{R}$ is a spatial regularization functional, e.g. a vectorial form of the total variation applied to the abundances A. Note that without the spatial regularization term, the problem becomes separable: a quadratic minimization over the simplex can be carried out for every pixel, independently. $\langle \cdot,\cdot \rangle$ is the usual Frobenius inner product. The elements of C are :

$$c_{m,n} = \parallel d_{\bullet,m} - y_{\bullet,n} \parallel^2 \tag{2}$$

$c_{m,n}$ is a measure of how far, spectrally, the pixel y_n is from the endmember d_m. Hence, $\langle A,C \rangle$ will tend to binarize the output : if $c_{m,n}$ is large, the abundance of d_m in the reconstruction will decrease progressively to 0, depending on the value of α. Parameter $\alpha \in [0,1]$ hence appears as a tuning parameter. Varying α leads to a variety of results progressively ranging from the standard LMM ($\alpha = 0$, low sparsity) to a hard classification ($\alpha = 1$, high level of sparsity: only one endmember is selected for each pixel). The optimal solution lies between these two extreme situations (see section 3).

While the first term in (1), along the spatial regularization, is the classical LMM, the second term can be understood as the convex relaxation of a hard classification. Indeed, hard classification aims at assigning only one endmember at every pixel:

$$\begin{aligned} &minimize_A \parallel DA - Y \parallel^2 + \mathcal{R}(A) \\ &s.t. \quad a_{\bullet,n} \in \mathcal{B}_M, \forall n = 1,\dots,N \end{aligned} \tag{3}$$

where $\mathcal{B}_M$ is the set of **binary vectors** of $\mathbb{R}_M$; that is, with only null elements, but one element equal to one. Note that without the spatial regularization term, the solution is trivial: pixel number n is represented by one single endmember, the endmember $d_{\bullet,m}$ (the m-th column of D) which minimizes (2).
However, when the spatial regularization term is introduced, the nonconvex problem (3) becomes NP-hard, so one wants to replace it by a convex problem. It turns out that considering the convex envelope of this problem consists in replacing the constraint of binarity by the constraint of belonging to the simplex. Consequently the convex envelope of (3) is not the linear model, but the following expression:

$$\begin{aligned} &minimize_A \langle A,C \rangle + \mathcal{R}(A) \\ &s.t. \quad a_{\bullet,n} \in \Delta_M, \forall n = 1,\dots,N \end{aligned} \tag{4}$$

using the previous definitions of $\langle A,C \rangle$ and C.
Indeed,

$$\begin{aligned} \parallel DA - Y \parallel^2 &= \sum_{n=1}^{N} \parallel Da_{\bullet,n} - y_{\bullet,n} \parallel^2 \\ &= \sum_{n=1}^{N} \parallel (\sum_{m=1}^{M} a_{m,n} d_{\bullet,m}) - y_{\bullet,n} \parallel^2 \end{aligned} \tag{5}$$

and

$$\langle A,C \rangle = \sum_{n=1}^{N} \sum_{m=1}^{M} a_{m,n} \parallel d_{\bullet,m} - y_{\bullet,n} \parallel^2 \tag{6}$$

and one can check that the cost functions in 5 and 6 are equal if $a_{\bullet,n}$ is binary for every n.

On the other hand, for $a_{\bullet,n}$ in the simplex instead of binary, the cost function in (5) takes a lower value than the one in (6), which is penalizing: one wants the cost function to take a large value for nonbinary abundance vectors, so that when the cost function is minimized, binary vectors are favored.

In order to solve the optimization problem (1), we propose to use the primal-dual algorithm presented by Condat in [4] and detailed in the following section.

2.2 Primal dual algorithm

We provide in this section some background information for the optimization. Let $\mathcal{H}$ be a Hilbert space with its inner product $\langle \cdot,\cdot \rangle$ and norm $\| \cdot \| = \langle \cdot,\cdot \rangle^{1/2}$. We denote by $\Gamma_0(\mathcal{H})$ the set of proper, lower semicontinuous, convex functions from $\mathcal{H}$ to $\mathbb{R} \cup +\infty$. Let $\mathcal{J}$ belong to $\Gamma_0(\mathcal{H})$. Its domain is $dom(\mathcal{J}) = s \in \mathcal{H} : \mathcal{J}(s) < +\infty$. Its Fenchel-Rockafellar conjugate $\mathcal{J}^* \in \Gamma_0(\mathcal{H})$ is defined by $\mathcal{J}^*(s) = sup_{s' \in \mathcal{H}}[\langle s,s' \rangle - \mathcal{J}(s')]$ and its proximity operator by
$prox_{\mathcal{J}}(s) = argmin_{s' \in \mathcal{H}}[\mathcal{J}(s') + \frac{1}{2} \parallel s - s' \parallel]$. The strong relative interior of a convex subset Ω of $\mathcal{H}$ is denoted by $sri(\Omega)$.

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Hilbert spaces. The method aims at solving the following primal problem :

$$Find \quad \hat{x} \in argmin_{x\in\mathcal{X}}[F(x) + G(x) + H(Lx)] \quad (7)$$

where:

- F: $\mathcal{X} \to \mathbb{R}$ is convex, differentiable on $\mathcal{X}$ and its gradient ∇F is β-Lipschitz continuous, for some $\beta \in [0,+\infty[$, that is :

$$\| \nabla F(x) - \nabla F(x') \| \leq \beta \| x - x' \|, \quad \forall (x,x') \in \mathcal{X}^2 \quad (8)$$

- $G \in \Gamma_0(\mathcal{X})$ and $H \in \Gamma_0(\mathcal{Y})$ have proximity operators which can be solved efficiently with high precision (or have a closed-form representation).
- $L : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator with adjoint L^*.
- The set of minimizers of (7) is supposed nonempty

The corresponding dual formulation of the primal problem (7) is :

$$Find \quad \hat{y} \in argmin_{y\in\mathcal{Y}}[(F+G)^*(-L^*y)+H^*(y)] \quad (9)$$

where $(F + G)^*(-L^*y)$ is an infimal convolution.

Another formulation of problems (7) and (9) is to combine them into the search of a saddle-point of the Lagrangian:

$$Find(\hat{x},\hat{y}) \in argmin_{x\in\mathcal{X}} max_{y\in dom(\mathcal{H}^*)}[F(x) + G(x) - H^*(y) + \langle Lx, y\rangle] \quad (10)$$

If the following holds:

$$0 \in sri\Big(L(dom(G)) - dom(H))\Big) \quad (11)$$

The set of solutions to (9) is nonempty, and for every primal solution $\hat{x}$ to (7) and solution $\hat{y}$ to (9), then $(\hat{x},\hat{y})$ is a solution to (10). This also holds if :

$$\frac{1}{\tau} - \sigma \| L \|^2 \geq \frac{\beta}{2} \quad (12)$$

where β is as defined in (8). Note that all of these requirements are satisfied for our problem.

2.3 Implementation

The spatial regularization functional in (1), $\mathcal{R}(A)$ used here was in the form of a total variation over the abundances A, with a tuning coefficient λ. For the primal-dual iterations, we set $\rho = 0.9 \quad \forall n \in \mathbb{N}$ (see Algorithm 1). We calculate the 2-norm of the spatial gradient operator L, and then set:

- $\tau = 10^{-2.5}$
- $\eta = 0.49$
- $\sigma = (\frac{1}{\tau} - \frac{\beta}{\eta})/ \| L \|_2^2$

so as to satisfy the conditions in [4]. We want to find the solution to (1). Let:

$$F(A) = (1-\alpha) \| DA - Y \|^2 + \alpha\langle A,C\rangle \quad (13)$$

then

$$\nabla F(A) = (1-\alpha)(Y - AD)D' + \alpha C \quad (14)$$

We minimize the following functional:

$$E(A) = F(A) + \delta_{\Delta_M}(A) + \lambda TV(A) \quad (15)$$

for some parameter $\alpha \in [0,1]$.

where the spatial regularization corresponds to $\mathcal{R}(A)$ in (1). δ_{Δ_M} is the indicator function on the $M - simplex$.The notation for $\delta_{\Delta_M}(A)$ in (15) is used to make the expression lighter, it is actually evaluated columnwise. We can then use the primal dual algorithm to solve the saddle-point problem described in [4]. The initialization doesn't impact the output (since the problem is convex), therefore we set $x_0 = 0$ and $y_0 = 0$. The *maximum* number of iterations is set to $maxiter = 300$, and a convergence criterion is set as $tol_A = 5 \times 10^{-4}$.

Algorithm 1 Primal-Dual iterations for hyperspectral images

Iterate, for every $0 \leq n \leq 300$
or while $A_{n+1} - A_n > tol$:

1: $\tilde{A}_{n+1} = prox_{\tau\delta_{\Delta_M}}(A_n - \tau(\nabla F(A_n) + e_{F,n}) - \tau L^* S_n) + e_{\delta_{\Delta_M},n})$
2: $\tilde{S}_{n+1} = prox_{\sigma H^*}(S_n - \sigma L(2\tilde{A}_{n+1} - A_n)) + e_{H,n})$
3: $(A_{n+1}, S_{n+1}) = \rho(\tilde{A}_n + 1, \tilde{S}_n + 1) + (1-\rho)(A_n, S_n)$

where the error terms $e_{F,n} \in \mathcal{X}, e_{G,n} \in \mathcal{X}$, $e_{H,n} \in \mathcal{Y}$ model the inexact computation of the operators $\nabla F, prox_{\tau G}, prox_{\sigma H^*}$, respectively.

Convergence is reached within 150 iterations and within 200s for large datasets (250x190x188), with average runtimes around 120-150s (MATLAB implementation).

For the Synthetic dataset, made up of 40,000 pixels and 162 bands (described in the next section), the runtime was within a minute for 6 endmembers.

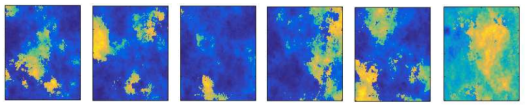
3 RESULTS

The algorithm was tested on several datasets, both synthetic and real.

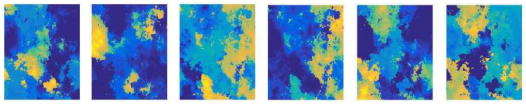
3.1 Synthetic Dataset

A synthetic dataset was created, using six endmembers. First, the abundance maps were generated as a mixture of Gaussians, each map containing 1 pure pixel, for five of the endmembers. The sum-to-one constraint over the abundances holds for these 5 endmembers. The sixth endmember was added in a second step, with an abundance ranging from 0 to 0.1. This additionnal endmember mimics common situations where the unmixing algorithm overfits the data and considers non-existing

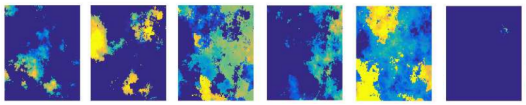
endmembers to optimize the reconstruction. Such endmembers usually come with small abundances. The abundances and endmembers were then mixed, without normalization (see figure 1). The ground truth for the figure is displayed in figure 1a. All the results are provided with $\lambda = 10^{-2}$.



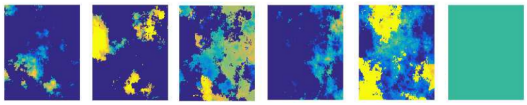
(a) Ground truth.



(b) $\alpha = 0$.



(c) $\alpha = 0.1$.



(d) $\alpha = 0.2$.

Figure 1: Recovered abundance maps for different values of α. Synthetic dataset with no normalization. The values range from 0 (blue) to 1 (yellow).

As α increases the abundances of the 6th endmember quickly drops to zero, which is the desired outcome. Naturally, as α increases, the error on the other abundances - corresponding to the "true" endmembers - also increases. Overall, there is a non-zero value of α which minimizes the Root-Mean-Squared-Error (RMSE) computed on the abundances:

$$RMSE = \sqrt{(\sum_{k=1}^{N} \frac{(A(k) - A_{true}(k))^2}{N})} \quad (16)$$

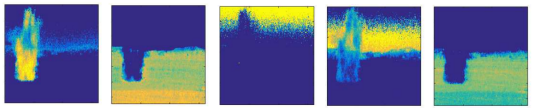
where A is the recovered abundance, A_{true} corresponds to the abundance of the ground truth and N is the number of pixels.

These results demonstrate the interest of the model (the optimal value does not correspond to any of the two extreme situations previously described). At this stage, determining the optimal value of α automatically is still an open problem, to be solved in future work. However, this paper provides the theoretical foundation of the method, and demonstrates the proof of concept. Parameter α can be tuned manually using a simple grid search. In our experiments, we found that setting the value of α is not critical, with a reasonable range of values leading to the optimal solution. We cannot present these extensive tests due to space limitations.

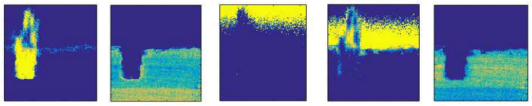
3.2 Real Dataset

The algorithm was also tested on real data, we present here a frame from a gas plume hyperspectral video sequence. The unmixing was performed with 5 endmembers, which is a reasonable value, as seen in [13],[18]. The results are shown for $\lambda = 10^{-2}$ (Figure 2). The endmembers are obtained using VCA.

Out of the 5 endmembers present in the scene, the endmember corresponding to the gas plume is the most impacted by the different values of α. The plume is diffuse and semi-transparent, and therefore the background endmembers (the mountains, ground and sky) all influence the unmixing of plume pixels. As was noted in [11], [13], [18] these are causes of nonlinearities. The horizontal trail observed is due to noise and corresponds to the mountain/sky interface. We can see that this error progressively disappears as α increases. However, the "hole" in the middle of the plume actually corresponds to gas that isn't as dense as its surroundings, but there still is gas present ([13]). As α increases, the hole becomes more important, and therefore our error increases.



(a) $\alpha = 0.3$



(b) $\alpha = 0.7$

Figure 2: Abundances for the plume dataset with different values of α and $\lambda = 10^{-2}$.

4 CONCLUSION

In this paper, we introduced a simple yet effective robust linear model and demonstrate its versatility. The theoretical foundation of the method and its solution using primal-dual optimization are discussed. The proof of concept is demonstrated by testing the method on both synthetic and real world datasets. The importance of errors in prior modeling, that are made when estimating the nature and number of endmembers, is diminished in the reconstruction, as are errors due to nonlinearities. Thanks to a very fast implementation, this technique is very competitive with respect to the regular linear model. Future work will have to provide theoretical bounds for the selection of the optimal value of α.

References

[1] Yoann Altmann, Abderrahim Halimi, Nicolas Dobigeon, and Jean-Yves Tourneret. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *Image Processing, IEEE Transactions on*, 21(6):3017–3025, 2012.

[2] José M Bioucas-Dias and José MP Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.

[3] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, 2012.

[4] Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2): 460–479, 2013.

[5] 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 Transactions on Image Processing*, 25(8):3890–3905, Aug 2016.

[6] L. Gao, L. Zhuang, and B. Zhang. Region-based estimate of endmember variances for hyperspectral image unmixing. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1807–1811, Dec 2016.

[7] A. Halimi, P. Honeine, and J. M. Bioucas-Dias. Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects. *IEEE Transactions on Image Processing*, 25(10):4565–4579, Oct 2016.

[8] Abderrahim Halimi, Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(11):4153–4162, 2011.

[9] Simon Henrot, Jocelyn Chanussot, and Christian Jutten. Dynamical spectral unmixing of multitemporal hyperspectral images. *IEEE Transactions on Image Processing*, 25(7):3219–3232, 2016.

[10] Rob Heylen and Paul Scheunders. A multilinear mixing model for nonlinear spectral unmixing. *Geoscience and Remote Sensing, IEEE Transactions on*, 54(1):240–251, 2016.

[11] Rob Heylen, Mario Parente, and Paul Gader. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 1844–1868, 2014.

[12] Jun Li and José M Bioucas-Dias. Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 3, pages III–250. IEEE, 2008.

[13] Ekaterina Merkurjev, Justin Sunu, and Andrea L Bertozzi. Graph mbo method for multiclass segmentation of hyperspectral stand-off detection video. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 689–693. IEEE, 2014.

[14] José MP Nascimento and José MB Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[15] Pierre-Antoine Thouvenin, Nicolas Dobigeon, and Jean-Yves Tourneret. Online unmixing of multitemporal hyperspectral images accounting for spectral variability. *IEEE Transactions on Image Processing*, 25(9):3979–3990, 2016.

[16] Gregg Vane, Robert O Green, Thomas G Chrien, Harry T Enmark, Earl G Hansen, and Wallace M Porter. The airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 44(2-3):127–143, 1993.

[17] M. A. Veganzones, G. Tochon, M. Dalla-Mura, A. J. Plaza, and J. Chanussot. Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *IEEE Transactions on Image Processing*, 23(8):3574–3589, Aug 2014.

[18] W. Zhu, V. Chayes, A. Tiard, S. Sanchez, D. Dahlberg, A. L. Bertozzi, S. Osher, D. Zosso, and D. Kuang. Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2786–2798, May 2017.