

Estimating the number of endmembers to use in spectral unmixing of hyperspectral data with collaborative sparsity

Lucas Drumetz¹, Guillaume Tochon², Jocelyn Chanussot¹, Christian Jutten^{1*}

¹Univ. Grenoble Alpes, CNRS, GIPSA-lab, F-38000 Grenoble, France
{lucas.drumetz,jocelyn.chanussot,christian.jutten}@gipsa-lab.fr,

²EPITA Research and Development Laboratory (LRDE), FR-94276 Le Kremlin-Bicêtre, France
guillaume.tochon@lrde.epita.fr,

Abstract. Spectral Unmixing (SU) in hyperspectral remote sensing aims at recovering the signatures of the pure materials in the scene (endmembers) and their abundances in each pixel of the image. The usual SU chain does not take spectral variability (SV) into account, and relies on the estimation of the Intrinsic Dimensionality (ID) of the data, related to the number of endmembers (NOE) to use. However, the ID can be significantly overestimated in difficult scenarios, and sometimes does not correspond to the desired scale and application dependent NOE. Spurious endmembers are then frequently extracted and included in the model. We propose an algorithm for SU incorporating SV, using collaborative sparsity to discard the least explicative endmembers in the whole image. We compute an algorithmic regularization path for this problem to select the optimal set of endmembers using a statistical criterion. Results on simulated and real data show the interest of the approach.

Keywords: Hyperspectral images, remote sensing, collaborative sparsity, Alternating Direction Method of Multipliers, regularization path, Bayesian Information Criterion

1 Introduction

The fine spectral resolution of hyperspectral remote sensing images allows to precisely identify and characterize the materials of the observed scene. However, this spectral resolution comes at the price of having a coarser spatial resolution than classical color or even multispectral images. Therefore, there are often several materials of interest present at the same time in the Field of View (FOV) of the sensor during the acquisition of a pixel, and the resulting observed spectrum is a mixture of the contributions of these materials. Spectral Unmixing (SU) is a (blind) source separation problem whose goal is to recover the spectral signatures of the pure materials in the scene (called *endmembers*) and to estimate their proportions in each pixel (called *fractional abundances*)[3]. To do that, a Linear Mixing Model (LMM) is often considered, assuming as a first approximation that the contributions of each endmember in each pixel sum up in a linear

* This work has been partially supported by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013, under Grant Agreement no.320684 (CHESS project)

way, with the abundances as weights. In order to interpret the abundances as fractions, they are usually constrained to be positive and to sum to one in each pixel. The classical linear SU unmixing chain is usually divided into three steps: i) Estimating the number of endmembers to use, which is a hard scale and application dependent (not to mention somewhat subjective) task, using intrinsic dimensionality (ID) estimation algorithms [2, 5], ii) Extracting the endmembers' signatures, generally with geometric approaches, such as the Vertex Component Analysis (VCA) [14] (usually assuming there are pure pixels in the image), and iii) Estimating the abundances by constrained least squares, using for instance the Fully Constrained Least Squares Unmixing (FCLSU) algorithm of [10]. The main two limitations of this strategy have been identified as nonlinearities [11] and spectral variability (see [6] and references therein). Nonlinear mixtures can occur when each ray of light received by the sensor has interacted with more than one material (e.g. in tree canopies, urban scenarios or in particulate media, such as sand). Dealing with spectral variability amounts to consider that each material possesses a certain intra-class variability, which is not what the usual approach does since it implicitly considers that every material is perfectly represented by a single spectral signature. If these two limitations of the usual SU chain are well identified and currently receiving a lot of attention, much less emphasis is put on the estimation of the number of endmembers to use. This number is often considered to be the same as the ID concept. If these two quantities indeed coincide when the LMM holds on simulated data, there is no such guarantee in nonlinear scenarios or when spectral variability is significant. In addition, the ID can be affected by outliers, which are usually not wanted in SU results. The ID has been shown to be subject to overestimation for several algorithms in difficult scenarios (small spatial dimensions, high spectral dimension, significant noise level) [7]. The errors committed at this step are then propagated to the whole unmixing chain, since spurious endmembers are extracted and incorporated to the model. In this paper, we propose an algorithm to perform linear SU of hyperspectral data, incorporating spectral variability, while automatically identifying the wrongly extracted endmembers and removing them from the pool of endmembers during the SU process, in order to keep the most relevant only. To this end, starting from the likely overestimated ID, we define an optimization problem using collaborative sparsity [13], so that irrelevant endmembers, usually associated with sparse and meaningless abundance maps, do not contribute in any pixel of the image. In order to select the appropriate number of endmembers to retain, we compute an algorithmic regularization path [12] for the optimization problem, providing a sequence of smaller and smaller candidate endmember matrices. The sequence goes from the whole initial pool of endmembers to a fully sparse model, each time removing the least explicative endmember in the current matrix. The only step remaining is to select the most appropriate element of this sequence for the problem at hand. We use the Bayesian Information Criterion (BIC) to select the optimal model, favoring models reconstructing the data well with a limited number of parameters.

The remainder of this paper is organized as follows: section 2 presents the pro-

posed approach in detail, then section 3 shows the results of experiments conducted on simulated and real data, and finally section 4 gathers some concluding remarks.

2 Proposed Approach

2.1 Extended Linear Mixing Model

Once the ID of the dataset has been estimated, one usually resorts to an end-member extraction algorithm, such as the VCA to obtain the spectra of the pure materials. The next step is the estimation of the abundances. For a hyperspectral image $\mathbf{X} \in \mathbb{R}^{L \times N}$, where L is the number of spectral bands, and N is the number of pixels, an endmember matrix $\mathbf{S}_0 \in \mathbb{R}^{L \times d}$ has been extracted (where d is the estimated ID). We denote the abundance matrix by $\mathbf{A} \in \mathbb{R}^{d \times N}$. In the usual linear SU setting, the abundances are estimated by nonnegative linear least squares, with the additional abundance sum to one constraint (ASC). However, here, following [17], we change the mixing model to incorporate SV into the unmixing at a negligible cost. We consider the following mixing model:

$$\mathbf{x}_k = \psi_k \sum_{p=1}^d a_{pk} \mathbf{s}_{0p} + \mathbf{e}_k, \quad (1)$$

where \mathbf{x}_k is the k^{th} column of \mathbf{X} , (i.e. the spectrum of pixel k), and \mathbf{s}_{0p} is the p^{th} column of \mathbf{S}_0 (i.e. the reference endmember for material p). \mathbf{e}_k is an additive noise, usually assumed to be Gaussian. a_{pk} is the abundance coefficient of endmember p in pixel k . Finally, ψ_k is a scaling factor, which models SV effects in each pixel, e.g. locally changing illumination conditions in the image, due to the topography of the imaged scene and to the photometric properties of the materials. This model is a simplified version of the Extended Linear Mixing Model (ELMM) [17, 8] (which considers distinct scaling factors for each material). With this model, \mathbf{S}_0 is then a reference endmember matrix, and we can define local endmember matrices in each pixel by computing $\mathbf{S}_k \triangleq \psi_k \mathbf{S}_0$. If this model holds, [17] shows that the quantity estimated by nonnegative least squares in each pixel and for each material actually incorporates SV information, via the product $\phi_{kp} \triangleq \psi_k a_{kp}$. The model reduces to the LMM when all scaling factors are equal to 1. CLSU (for Constrained Least Squares Unmixing) solves, for each pixel:

$$\arg \min_{\phi_k \geq \mathbf{0}} \frac{1}{2} \|\mathbf{x}_k - \mathbf{S}_0 \phi_k\|_2^2, \quad (2)$$

where $\phi_k \in \mathbb{R}^d$ collects all the ϕ_{kp} for a given pixel k . If we sum all the entries of ϕ_k , we obtain:

$$\sum_{p=1}^d \phi_{pk} = \sum_{p=1}^d \psi_k a_{kp} = \psi_k \sum_{p=1}^d a_{kp} = \psi_k, \quad (3)$$

by reintroducing the ASC on the *actual* abundances, and not their product with the scaling factor. Then we can easily obtain $\mathbf{a}_k = \frac{\phi_k}{\psi_k}$.

2.2 Collaborative sparsity for hyperspectral unmixing

Our goal is to eliminate the wrongly estimated endmembers from the SU process. To do that, we use collaborative sparsity [13]. This concept, also known

as Multiple Measurement Vector (MMV), or joint sparsity in the signal processing community, extends regular sparsity to a collection of signals which are encouraged to share the same support. For our application, we would like the abundances of the least explicative endmembers (and hence their product with the scaling factors) to be zero on the whole image support. This can be done by considering the following optimization problem:

$$\begin{aligned} & \arg \min_{\Phi \geq 0} \|\Phi\|_{\text{row},0} \\ & \text{s.t. } \|\mathbf{X} - \mathbf{S}_0\Phi\|_F^2 < \delta \end{aligned} \quad (4)$$

where $\|\cdot\|_{\text{row},0}$ is the row-wise \mathcal{L}_0 norm (computing the number of nonzero rows of its matrix argument) of the whole matrix $\Phi \in \mathbb{R}^{d \times N}$, $\|\cdot\|_F$ is the Frobenius norm, and δ is a desired data fit value. This problem allows us to discard entire rows of the feature matrix Φ , but is nonconvex, combinatorial and NP-hard. In order to obtain a more friendly formulation, we consider the following convex relaxation:

$$\arg \min_{\Phi} \frac{1}{2} \|\mathbf{X} - \mathbf{S}_0\Phi\|_F^2 + \lambda \|\Phi\|_{2,1} + \mathcal{I}_{\mathbb{R}_+^{d \times N}}(\Phi), \quad (5)$$

where $\mathcal{I}_{\mathbb{R}_+^{d \times N}}$ is the indicator function of the positive orthant of $\mathbb{R}^{d \times N}$, and λ is a regularization parameter. The quantity $\|\cdot\|_{2,1}$ is the mixed $\mathcal{L}_{2,1}$ norm, defined for any matrix $\Phi \in \mathbb{R}^{d \times N}$ as:

$$\|\Phi\|_{2,1} = \sum_{i=1}^d \left(\sum_{j=1}^N |\phi_{ij}|^2 \right)^{\frac{1}{2}} = \sum_{i=1}^d \|\phi_i\|_2, \quad (6)$$

where ϕ_i is the i^{th} row of Φ .

The $\mathcal{L}_{2,1}$ norm encourages row-wise sparsity in the feature matrix, because it is the \mathcal{L}_1 norm of a vector made of the \mathcal{L}_2 norms of the rows of this matrix. Consequently, many of these \mathcal{L}_2 norms will be zero (or close to zero), which will produce the desired effect of nulling the coefficients of irrelevant endmembers in all pixels. This problem can be readily solved using proximal algorithms, for instance the Alternating Direction Method of Multipliers (ADMM) [4]. To use it, we introduce split variables to decouple the different terms in the optimization. We then rewrite problem (5) in an equivalent formulation using linear constraints, which are suitable for the ADMM:

$$\begin{aligned} & \arg \min_{\Phi} \frac{1}{2} \|\mathbf{X} - \mathbf{S}_0\Phi\|_F^2 + \lambda \|\mathbf{U}\|_{2,1} + \mathcal{I}_{\mathbb{R}_+^{d \times N}}(\mathbf{V}) \\ & \text{s.t. } \mathbf{U} = \Phi, \mathbf{V} = \Phi. \end{aligned} \quad (7)$$

The ADMM uses an Augmented Lagrangian (AL) approach to split the hard nondifferentiable problem of Eq. (5) into several easier subproblems w.r.t. the two blocks of variables \mathbf{U} and \mathbf{V} , and a so-called *dual update* of the introduced Lagrange multipliers (called \mathbf{C} and \mathbf{D} below), all with closed form solutions, which can be iterated until convergence. Collaborative sparsity then seems like a good candidate to discard the unwanted spurious endmembers. However, there are two problems with this approach. The first is that since the linear constraints of the ADMM are only satisfied asymptotically, we have no guarantee that all the

entries of the supposedly discarded rows of the feature matrix Φ will be exactly zero (and this actually happens in practice). Then an arbitrary thresholding step is required to eliminate endmembers with a small contribution [13, 1]. The second is that in order to obtain the appropriate sparsity level, the regularization parameter λ needs to be optimized through a grid search, which is computationally very costly, and requires a criterion to select the best run of the algorithm. The next section provides solutions for both issues.

2.3 Computing a regularization path

In order to tackle both the regularization parameter issue and the inexact sparsity of the collaborative sparse regression at once, we would like to obtain the regularization path of the solution, as a function of λ . Regularization paths can sometimes be computed cheaply, for instance on the LASSO (for Least Absolute Shrinkage Selection Operator) problem [9]. However, for more complex problems, such as ours, there is no way, to our knowledge, to obtain this regularization path easily. A convenient workaround for this is to compute a so-called ADMM algorithmic regularization path, introduced in [12]. This approach is able to use the ADMM to quickly approximate the sequence of active supports of the variable of interest, when the regularization parameter increases, for certain sparsity regularized least squares problems. Even though there are as of today no theoretical guarantees on the efficiency of this algorithm, it was experimentally shown to be able to efficiently approximate the true sequence of active sets on several problems [12], including the LASSO. Here, we extend this algorithm to collaborative sparsity.

Since exactly solving the optimization problem for a large number of regularization parameters would be too time consuming, we are more interested in finding the active set of endmembers when the weight of the sparsity term increases w.r.t. this of the data fit term. The idea is to find a sequence of endmember matrices, whose number of endmembers is decreasing from d to zero (when the model is fully sparse). Each new matrix contains the same endmembers as the previous one, except for the one (or the few ones) which is going to be discarded next, when the weight of the sparsity term gets more important. We modify the ADMM in order to quickly obtain the support of the regularization path. An iteration of the ADMM is carried out for a very small value of the regularization parameter (guaranteeing a fully dense solution). Then, the variables obtained after this iteration are used as a warm start for another iteration with a new slightly higher regularization parameter. By repeating this for several iterations with higher and higher regularization parameters, the split variable \mathbf{U} , which undergoes a block soft thresholding (the proximal operator of the $\mathcal{L}_{2,1}$ norm [4]) becomes increasingly sparse. Since we are using warm starts, and because regularization parameters vary slowly, even if the ADMM is not fully converged at each iteration, the support of the active set is encoded in \mathbf{U} , often in one iteration only, long before this active set is propagated to Φ (this will be the case only at convergence, when the constraints of problem (7) are satisfied). With these modifications, we obtain Algorithm 1. ρ is the barrier parameter of the ADMM, which we fix to 1 throughout the paper, so that it does not interfere with the

tuning of the regularization parameter. \mathbf{soft}_τ denotes the block soft thresholding operator with scale parameter in index. If $\mathbf{u} \in \mathbb{R}^N$, $\mathbf{soft}_\tau(\mathbf{u}) = (1 - \frac{\tau}{\|\mathbf{u}\|_2})_+ \mathbf{u}$, where $\cdot_+ = \max(\cdot, 0)$ (and we have $\mathbf{soft}_\tau(\mathbf{0}) = \mathbf{0}$). This operator is applied row-wise to a matrix. $\mathbf{proj}_{\mathbb{R}_+^{d \times N}}$ denotes the projection onto the positive orthant (a simple thresholding). Here, we are using a geometric progression for γ (previously denoted as λ^1), whose common ratio is t . This value should be small to approximate the active sets of the regularization path well enough. The regularization space can be explored very quickly since the algorithm provides around d endmember subsets of the full endmember set extracted by VCA, that need to be tested after this process.

Data: \mathbf{X}, \mathbf{S}_0

Result: The sequence of \mathbf{U}^i , $i = 0, \dots, i_{max}$

Initialize Φ^0 and choose γ^0 and $t > 0$;

while $\|\mathbf{U}^i\|_{row,0} \neq 0$ **do**

$\gamma^i \leftarrow t\gamma^{i-1}$;

$\mathbf{U}^i \leftarrow \mathbf{soft}_{\gamma^i/\rho}(\Phi^{i-1} - \mathbf{C}^{i-1})$;

$\Phi^i \leftarrow (\mathbf{S}_0^\top \mathbf{S}_0 + 2\rho \mathbf{I}_d)^{-1} (\mathbf{S}_0^\top \mathbf{X} + \rho(\mathbf{U}^i + \mathbf{V}^{i-1} + \mathbf{C}^{i-1} + \mathbf{D}^{i-1}))$;

$\mathbf{V}^i \leftarrow \mathbf{proj}_{\mathbb{R}_+^{d \times N}}(\Phi^i - \mathbf{D}^{i-1})$;

$\mathbf{C}^i \leftarrow \mathbf{C}^{i-1} + \mathbf{U}^i - \Phi^i$;

$\mathbf{D}^i \leftarrow \mathbf{D}^{i-1} + \mathbf{V}^i - \Phi^i$;

$i \leftarrow i + 1$

end

Algorithm 1: ADMM algorithmic regularization path for problem (7).

2.4 Selecting the best model

Using the active sets \mathbf{U}^i , we can recover a sequence of sparser and sparser candidate endmember matrices (whose i^{th} element is denoted as \mathbf{S}_0^i). The last step is to select the optimal endmember matrix in the sense of some criterion. We use the Bayesian Information Criterion (BIC) [16], which helps choosing from a set of candidate models, by favoring those with an important likelihood, and penalizing those with a high number of parameters. This criterion assumes that the noise is Gaussian, spectrally and spatially white, a strong but still widely used assumption. A candidate model M_i is made of one of the \mathbf{S}_0^i and the corresponding estimated feature matrix with CLSU. For our problem, the BIC writes [15]:

$$BIC_i = \ln(L)P_i + L \ln \left(\frac{\|\mathbf{X} - \mathbf{S}_0^i \hat{\Phi}^i\|_F^2}{L} \right), \quad (8)$$

where P_i is the number of endmembers in $\mathbf{S}_0^i \in \mathbb{R}^{L \times P_i}$. $\hat{\Phi}^i$ is the abundance matrix estimated by CLSU using the data and the endmember matrix \mathbf{S}_0^i . The best model is simply the one minimizing the BIC value, providing $P \leq d$ endmembers and abundance maps. The endmembers which do not contribute much to the data fit have been discarded.

3 Results

In this section, we show the results of the proposed approach on a simulated and a real dataset. We compare the obtained results to those of the classical

¹ We changed the notation of this regularization parameter, because we do not completely solve the optimization problem (7).

SU chain, using the Hyperspectral Subspace Identification by Minimum Error (HySIME) algorithm [2] for ID estimation, VCA for endmember extraction, and nonnegative least squares and the normalization detailed in section 2.1 for the abundance and scaling factor estimation, so that the results follow the ELMM. We call this approach HySIME + S-CLSU (for Scaled CLSU).

3.1 Results on simulated data

For the simulated data, we voluntary put ourselves in a case where ID estimation algorithms are prone to overestimation, namely an image with small spatial dimensions, a high spectral dimension, and non negligible noise values [7]. 6 endmembers were randomly selected from the United States Geological Survey (USGS) spectral library, containing in-situ acquired spectra of various minerals. We built synthetic abundance maps for 6 materials using Gaussian Random Fields. We also computed scaling factor maps for each material using mixtures of Gaussians. These two quantities are shown in Fig. 1. Since the actual contribution of a material to a pixel is the product between the abundance and the scaling factor, the effect of either quantity will only be noticeable when the other is sufficient. For example, significant SV for a material with a small abundance will be very hard to recover. We mixed the data using the (full) ELMM of [17, 8] and finally added white Gaussian noise, such that the SNR was 25 dB. This resulted in a $40 \times 40 \times 300$ hyperspectral image. The HySIME algo-

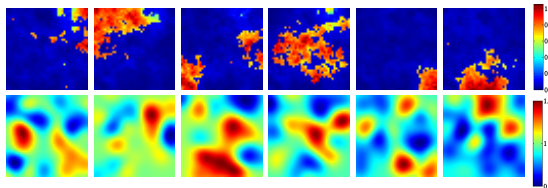


Fig. 1. True abundance (top row) and scaling factor (bottom row) maps for the synthetic data.

rithm estimated the ID of this dataset to be 16, whereas only 6 endmembers were used in the data generation (as a comparison, the Random Matrix Theory (RMT) based algorithm of [5] returned an ID value of 25). We show the abundance estimation results for our approach and the classical one in Fig. 2. We can see that the HySIME + S-CLSU approach is able to recover correctly 4 of the 6 abundance maps. However, the last two abundances are split in 11 different maps, which correspond to unnecessary (because very close between them) signatures extracted by the VCA. The proposed approach (with the parameters empirically tuned to $\gamma^0 = 10^{-4}$ and $t = 1.01$) only retained 6 endmembers in that case, all of which are associated with abundance maps which are very close to the real ones. We do not show the scaling factor maps here for lack of space, but they are very similar in both cases, and each pixel value accounts for the scaling factor of the predominant material. In a scenario where the full ELMM would be used, the scaling factors would probably be much easier to interpret for our approach, because we would be able to distinguish the contributions of

each material, and better explain what happens in heavily mixed pixels. Fig. 3 shows the BIC values we get for the obtained sequence of candidate endmember matrices. While the data fit term decreases continually every time we add an endmember, the decrease is marginal after 6 endmembers, while the number of parameters is more and more penalized. The BIC then reaches a minimum for 6 endmembers. This shows that the unnecessary endmembers only fitted the noise, while being difficult to interpret. For each material, we also computed a Root Mean Squared Error (RMSE) on the abundances for material p , with $aRMSE = \frac{1}{N} \sum_{k=1}^N \|\mathbf{a}_{\text{true},pk} - \hat{\mathbf{a}}_{pk}\|_2$, with $\hat{\mathbf{a}}_{pk}$ the closest abundance map to the true one in the results of one of the two algorithms (see Table 1). We see that except for material 2 (the materials are numbered from left to right in Fig. 1), where the proposed approach recovers a slightly noisy abundance map, the abundances are better recovered by the proposed method.

Material	1	2	3	4	5	6
Proposed method	1.0×10^{-3}	1.1×10^{-3}	5.9×10^{-4}	9.2×10^{-4}	4.2×10^{-4}	5.5×10^{-4}
HySIME + S-CLSU	4.6×10^{-3}	8.9×10^{-3}	7.3×10^{-4}	1.4×10^{-3}	5.2×10^{-4}	7.5×10^{-4}

Table 1. aRMSE values between the true abundance maps and the closest one of the two competing approaches. The best value is in red.

3.2 Results on real data

To confirm the soundness of our approach, we apply it to a 100×100 subset of the Washington DC mall dataset, acquired over the National Gallery of Art (shown in Fig. 4, with the endmembers extracted by VCA displayed as red crosses) by the HYDICE sensor, comprising 191 spectral bands in the visible and near infrared, with a spatial resolution of 2.8 m. HySIME estimated the ID to be 38 (the RMT algorithm returned 65). We show the estimated abundance maps in Fig. 5. We see that the abundance maps on the left are very hard to interpret, because many are very sparse and related to outlier pixels, while the proposed approach allowed to retain only the most important. Only one visually relevant endmember retained by HySIME+S-CLSU is not present in the proposed approach, corresponding to marble (stairs and dome of the museum). The scaling factor maps, shown in Fig. 4, are relatively similar in both cases, except in the grass part, where some

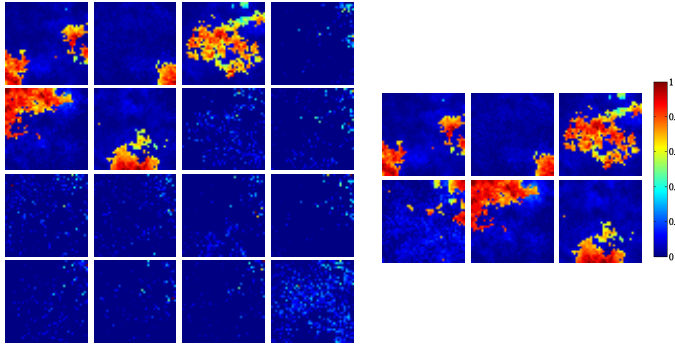


Fig. 2. Abundance maps extracted by HySIME + S-CLSU for the synthetic data (left), and by the proposed approach (right).

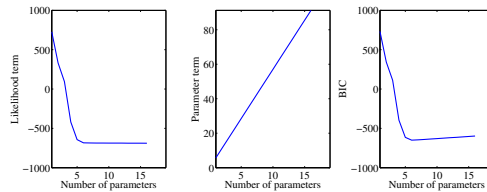


Fig. 3. Data fit term of the BIC (left), parameter term (middle) and BIC value (right), for the simulated data.

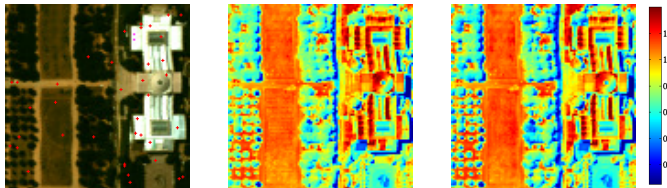


Fig. 4. RGB composition of the real dataset used (left), with the endmembers extracted by VCA in red crosses, and extracted scaling factor maps for Hysime + S-CLSU (middle) and the proposed approach (right).

geometrical structures appear for HySIME+S-CLSU, which could correspond to artifacts captured as one of the endmembers. Otherwise, the maps explain well the variability of the scene, with low values for the shadowed trees and structures, and higher for parts of the roofs exposed to the sun, for example.

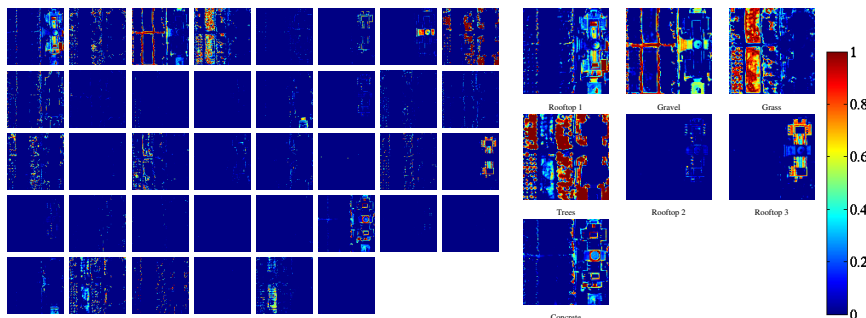


Fig. 5. Abundance maps extracted by HySIME + S-CLSU for the real data (left), and by the proposed approach (right).

4 Conclusion

We have presented a technique to overcome the likely overestimation of the number of endmembers to use in spectral unmixing of hyperspectral data, accounting for spectral variability. It is based on computing an approximate regularization path for a collaborative sparse regression problem, which allows to select the most relevant endmember signatures, and to discard the spurious ones. We confirmed the interest of the proposed approach on a synthetic dataset and a real one. Future work will include the full ELM (one scaling factor per material) to the framework of the proposed approach.

Bibliography

- [1] Ammanouil R, Ferrari A, Richard C, Mary D (2014) Blind and fully constrained unmixing of hyperspectral images. *IEEE Transactions on Image Processing* 23(12):5510–5518
- [2] Bioucas-Dias J, Nascimento J (2008) Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing* 46(8):2435–2445
- [3] Bioucas-Dias J, Plaza A, Dobigeon N, Parente M, Du Q, Gader P, Chanussot J (2012) Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(2):354–379
- [4] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1):1–122
- [5] Cawse-Nicholson K, Damelin S, Robin A, Sears M (2013) Determining the intrinsic dimension of a hyperspectral image using random matrix theory. *IEEE Transactions on Image Processing* 22(4):1301–1310
- [6] Drumetz L, Chanussot J, Jutten C (2016) Endmember variability in spectral unmixing: recent advances. In: *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp 1–4
- [7] Drumetz L, Veganzones MA, Gómez RM, Tochon G, Dalla Mura M, Licciardi GA, Jutten C, Chanussot J (2016) Hyperspectral local intrinsic dimensionality. *IEEE Transactions on Geoscience and Remote Sensing* 54(7):4063–4078
- [8] Drumetz L, Veganzones MA, Henrot S, Phlypo R, Chanussot J, Jutten C (2016) Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Transactions on Image Processing* 25(8):3890–3905
- [9] Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. *The Annals of statistics* 32(2):407–499
- [10] Heinz D, Chang CI (2001) Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39(3):529–545
- [11] Heylen R, Parente M, Gader P (2014) A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6):1844–1868
- [12] Hu Y, Chi E, Allen GI (2015) ADMM algorithmic regularization paths for sparse statistical machine learning. *arXiv preprint arXiv:150406637*
- [13] Iordache MD, Bioucas-Dias JM, Plaza A (2014) Collaborative sparse regression for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing* 52(1):341–354
- [14] Nascimento J, Bioucas Dias J (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 43(4):898–910
- [15] Priestley MB (1981) *Spectral analysis and time series*. Academic press
- [16] Schwarz G (1978) Estimating the dimension of a model. *Ann Statist* 6(2):461–464
- [17] Veganzones MA, Drumetz L, Marrero R, Tochon G, Dalla Mura M, Plaza A, Bioucas-Dias J, Chanussot J (2014) A new extended linear mixing model to address spectral variability. In: *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*