

UCLA

UCLA Previously Published Works

Title

Hyperspectral Image Unmixing With Endmember Bundles and Group Sparsity Inducing Mixed Norms

Permalink

<https://escholarship.org/uc/item/4b86v38m>

Journal

IEEE Transactions on Image Processing, 28(7)

ISSN

1057-7149

Authors

Drumetz, Lucas
Meyer, Travis R
Chanussot, Jocelyn
[et al.](#)

Publication Date

2019-07-01

DOI

10.1109/tip.2019.2897254

Peer reviewed

Hyperspectral Image Unmixing with Endmember Bundles and Group Sparsity Inducing Mixed Norms

Lucas Drumetz, *Member, IEEE*, Travis R. Meyer, Jocelyn Chanussot, *Fellow, IEEE*, Andrea L. Bertozzi, and Christian Jutten, *Fellow, IEEE*

Abstract—Hyperspectral images provide much more information than conventional imaging techniques, allowing a precise identification of the materials in the observed scene, but because of the limited spatial resolution, the observations are usually mixtures of the contributions of several materials. The spectral unmixing problem aims at recovering the spectra of the pure materials of the scene (endmembers), along with their proportions (abundances) in each pixel. In order to deal with the intra-class variability of the materials and the induced spectral variability of the endmembers, several spectra per material, constituting endmember bundles, can be considered. However, the usual abundance estimation techniques do not take advantage of the particular structure of these bundles, organized into groups of spectra. In this paper, we propose to use group sparsity by introducing mixed norms in the abundance estimation optimization problem. In particular, we propose a new penalty which simultaneously enforces group and within group sparsity, to the cost of being nonconvex. All the proposed penalties are compatible with the abundance sum-to-one constraint, which is not the case with traditional sparse regression. We show on simulated and real datasets that well chosen penalties can significantly improve the unmixing performance compared to the naive bundle approach.

Index Terms—Hyperspectral imaging, remote sensing, spectral unmixing, endmember variability, group sparsity, convex optimization

I. INTRODUCTION

HYPERSPECTRAL imaging, also known as imaging spectroscopy, is a technique which allows to acquire information in each pixel under the form of a spectrum of reflectance or radiance values for many –typically hundreds of– narrow and contiguous wavelengths of the electromagnetic spectrum, usually (but not exclusively) in the visible and infra-red domains [1]. The fine spectral resolution of these images allow an accurate identification of the materials present in the scene, since two materials usually have distinct spectral profiles. However, this identification is made harder by the

relatively low spatial resolution (significantly lower than panchromatic, color or even multispectral images). Therefore, many pixels are acquired with several materials in the field of view of the sensor, and the resulting observed signature is a mixture of the contributions of these materials. Spectral Unmixing (SU) is then a source separation problem whose goal is to recover the signatures of the pure materials of the scene (called *endmembers*), and to estimate their relative proportions (called *fractional abundances*) in each pixel of the image [2].

Usually, since the abundances are meant to be interpreted as proportions, they are required to be positive and to sum to one in each pixel. Another classical assumption made for SU is that the mixture of the contributions of the materials is linear, leading to the standard Linear Mixing Model (LMM) [3], [4]. Let us denote a hyperspectral image by $\mathbf{X} \in \mathbb{R}^{L \times N}$, gathering the pixels $\mathbf{x}_k \in \mathbb{R}^L$ ($k = 1, \dots, N$) in its columns, where L is the number of spectral bands, and N is the number of pixels in the image. The signatures \mathbf{s}_p , $p = 1, \dots, P$ of the P endmembers considered for the unmixing are gathered in the columns of a matrix $\mathbf{S} \in \mathbb{R}^{L \times P}$. The abundance coefficients a_{pk} for each pixel $k = 1, \dots, N$ and material $p = 1, \dots, P$ are stored in the matrix $\mathbf{A} \in \mathbb{R}^{P \times N}$. With these notations, the LMM writes:

$$\mathbf{x}_k = \sum_{p=1}^P a_{pk} \mathbf{s}_p + \mathbf{e}_k, \quad (1)$$

where \mathbf{e}_k is an additive noise (usually assumed to be Gaussian distributed). Eq. (1) can be rewritten in a matrix form for the whole image:

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{E}, \quad (2)$$

where $\mathbf{E} \in \mathbb{R}^{L \times N}$ comprises all the noise values. We keep in mind the constraints on the abundances: Abundance Nonnegativity Constraint (ANC) $a_{pk} \geq 0 \forall (p, k)$ and the Abundance Sum-to-one Constraint (ASC) $\sum_{p=1}^P a_{pk} = 1, \forall k$. This model is considered physically realistic when each ray of light reaching the sensor has interacted with no more than one material on the ground (the so-called checkerboard configuration) [3]. A nice property of the LMM is that, combined with the constraints on the abundances, it provides a strong geometrical structure to the problem: indeed the ANC and ASC force the abundances to lie in the probability simplex with P vertices, which we denote by Δ_P . Since the data are obtained via a linear transformation of the abundances, they are also constrained to lie in a simplex, whose vertices are the endmembers. The actual signal subspace is then

L. Drumetz is with IMT Atlantique, Lab-STICC, UBL, Technopôle Brest-Iroise CS 83818, 29238 Brest Cedex 3, France (e-mail: lucas.drumetz@imt-atlantique.fr)

T. R. Meyer and A.L. Bertozzi are with the Department of Mathematics at the University of California, Los Angeles, 90095 Los Angeles, CA, USA (e-mail: {tmeyer;bertozzi}@math.ucla.edu).

J. Chanussot and C. Jutten are Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France. * Institute of Engineering Univ. Grenoble Alpes (e-mail: {jocelyn.chanussot,christian.jutten}@gipsa-lab.grenoble-inp.fr).

This work was 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), as well as the Agence Nationale de la Recherche and the Direction Générale de l'Armement, under the project ANR-DGA APHYPIIS, by CNRS PICS 263484, and by NSF grant DMS-1118971, NSF grant DMS-1417674, ONR grant N000141210838, UC Lab Fees Research grant 12-LR- 236660. L. Drumetz was also supported by a Campus France outgoing postdoctoral mobility grant, PRESTIGE-2016-4 0006.

a $P - 1$ -dimensional subspace, embedded in the ambient space.

The typical blind SU chain comprises three steps:

- Estimating the number of endmembers to consider. This is a very hard and ill-posed problem in itself (because there is no such thing as an optimal number of endmembers in real data, among other reasons) and many algorithms have been considered in the community to try to obtain a good estimate [5].
- Extracting the spectra corresponding to the endmembers, a procedure referred to as endmember extraction. Then again, many Endmember Extraction Algorithms (EEAs) exist in the literature to tackle this problem, with various assumptions, the main one being the presence in the data of pure pixels, i.e. pixels in which only one material of interest is present [6]. These algorithms try to exploit the geometry of the problem by looking for extreme pixels in the data, which are the endmembers if the LMM holds. A popular EEA using the pure pixel assumption is the Vertex Component Analysis (VCA) [7].
- Finally, estimating the abundances using the data and the extracted endmembers. This step is usually carried out by solving a constrained optimization problem:

$$\arg \min_{\mathbf{A} \geq \mathbf{0}, \mathbf{1}_P^T \mathbf{A} = \mathbf{1}_N^T} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2, \quad (3)$$

where $\mathbf{1}$ denotes a vector of ones whose size is given in index, and $\|\cdot\|_F$ is the Frobenius norm. Solving this problem is often referred to as Fully Constrained Least Squares Unmixing (FCLSU) [8].

This unmixing chain is now standard and has been used with success over the last two decades. However, the main two limitations of this approach have been identified as nonlinear effects, and endmember variability. Nonlinear interactions may occur if light reaching the sensor has interacted with more than one material on the ground (e.g. in tree canopies, or in particulate media). These physical considerations have spanned a number of approaches taking some nonlinear effects into consideration; see [9], [10] for reviews on that subject. The other limitation, termed spectral variability, is simply based on the consideration that all materials have a certain intra-class variability, and therefore cannot be represented by a single signature, as in the conventional LMM [11], [12]. Note that even though both phenomena are usually considered separately, some approaches attempt to address both simultaneously [13]. Recently, a formal link between models addressing these two limitations was derived in [14].

Causes of variability of the materials from one pixel to the other include changing illumination conditions locally in the scene, especially because of a nonflat topography, which will change the local incident angle of the light and the viewing angle of the sensor. Also, physico-chemical changes in the composition of the materials induce modifications on their spectra. For example, the concentration of chlorophyll in grass or soil moisture content can change the signatures of these two materials. Several approaches have been designed for variability aware unmixing, that is to correct the abundances

due to the intra-class variations of the endmembers. The existing methods could be basically divided into two groups: dictionary or bundle-based approaches [15]–[17], which try to model an endmember by a certain number of instances of each material, and model-based approaches, which define a specific model for the variation of the endmembers, be it computational [18]–[20] or more physics-inspired [21]. Even though most methods dealing with spectral variability in the spatial domain, within a single image, studies dealing with temporal endmember variability have also been recently developed [22], [23].

This paper focuses on variability in the spatial domain using bundle-based approaches. Even though they lack the interpretability of their model based counterparts, they can still fit the framework of the LMM, while correcting the abundances with less assumptions on the data. The idea behind spectral bundles is to find a way to extract a dictionary of different instances of the materials directly from the data. Then one can simply replace the matrix \mathbf{S} in Eq. (2) by a dictionary made of spectral bundles $\mathbf{B} \in \mathbb{R}^{L \times Q}$, where $Q > P$ is the total number of endmember candidates. Since this dictionary (and the abundance coefficients as well) is organized into P groups, and since Q can be relatively large, it makes sense to consider that only a few atoms of the bundle dictionary are going to be used in each pixel. Moreover, a now common assumption in SU is that a few materials are active in each pixel, out of the P considered materials. These two assumptions, though similar, are not equivalent. In any case, they justify the use of sparsity in the SU problem. For instance, sparse regression using libraries of endmembers (i.e. semi-supervised unmixing) has been investigated in [24] and exploited by many authors thereafter. It is interesting to note that sparse unmixing, when performed using convex optimization, and more precisely \mathcal{L}_1 norm regularization, is not compatible with the ASC. We will come back to this issue in the next section. When dealing with coefficients organized in groups, a small number of which are supposed to be active, one may resort to approaches similar to the so-called group LASSO (for Least Absolute Shrinkage and Selection Operator) [25]. In SU, both approaches have been combined in semi-supervised unmixing in [26], when the dictionary can be organized into clusters of signatures.

The contributions of this paper are threefold. We first show that when using the classical FCLSU to estimate the abundances of the materials in the context of endmember bundles, it is possible to recover pixel-wise endmembers representing variability effects more subtly than with the signatures of the bundle dictionary, and we provide a simple geometrical interpretation for them. Second, further elaborating on the considerations and results of the conference paper [27], we improve the abundance estimation when bundles are used by considering three “social” sparsity inducing penalties [28], [29], whose interest we show by comparing their performance on simulated and real data to the classical naive approach. Third, while two of those penalties have already been used in various signal processing problems, the last one (a mixed

fractional $\mathcal{L}_{1, \frac{a}{b}}$ quasinorm with a, b integers) is new to the best of our knowledge, and presents remarkable properties, while raising interesting technical challenges in its optimization.

The remainder of this paper is organized as follows: Section II presents in more detail the basic procedure to obtain a bundle dictionary from the image data, based on [16]. While this is not the core of our method, it still serves as a basis for blind SU with bundles, and we present it here for the reader’s convenience. We also propose a new geometric interpretation to SU using spectral bundles with FCLSU. Section III introduces the proposed sparsity inducing penalties, and details how to handle the resulting optimization problem for abundance estimation in each case. Section IV tests the different penalties on synthetic and real data, and compares them to the standard bundle unmixing approach. Finally, Section V gathers some concluding remarks and possible future research avenues.

II. DEALING WITH SPECTRAL VARIABILITY WITH ENDMEMBER BUNDLES

A. Automated Endmember Bundles

Extracting endmember bundles from the image data is a procedure whose goal is to obtain various instances of each endmember, still in a blind unmixing setting. We present and will use the so-called Automated Endmember Bundles (AEB) approach of [16]. It is conceptually very simple and based on the following idea: since running a classical EEA once on the data provides the extreme points of the dataset, several instances of each material could be obtained by running an EEA on several subsets of the data. Then, one can randomly sample m subsets of the data of size n (pixels), and run an EEA on each subset in order to obtain P endmember candidates, and therefore $Q = mP$ signatures in total. It is possible that for a given run, an EEA extract several signatures corresponding to the same material. Thus, it is not guaranteed that each endmember is represented by exactly m instances. Some rare endmembers could also appear. Usually, however, the number of desired bundles is set as the estimated number of endmember for the whole image. In addition, even in a perfect case, the instances for each endmember are not aligned, i.e. the ordering of the extracted endmembers in different subsets is not the same, due to the stochasticity of the EEAs. Thus, a clustering step is required to group the signatures into bundles. In [16], the clustering is performed using the k-means algorithm, with the spectral angle as a similarity measure. The spectral angle has the nice property of being insensitive to scalings of the input vectors, which is a well-known effect of spectral variability related to changing illumination conditions [21].

This idea of extracting several endmember candidates in different parts of the image is related to the recent Local Spectral Unmixing approaches, which handle variability by assuming it is less important in *local* (as opposed to *random* here) subsets of the image. In this type of methods, the whole unmixing process is entirely carried out in each (connected) subset. Some examples of local unmixing include [30]–[33].

The main difference here is that we extract endmembers in random subsets, before gathering all those signatures into a common pool organized into groups, from which we estimate abundances for the whole image, not in local subsets.

In any case, the result of the clustering is a dictionary \mathbf{B} of endmember candidates. The clustering step defines a group structure¹ on the abundance coefficients and on the dictionary. We denote this group structure by G , and each group G_i , $i = 1, \dots, P$ contains m_{G_i} signatures, so that representative j of group G_i in the dictionary is denoted as $\mathbf{b}_{G_i, j}$. Then there are $Q = \sum_{i=1}^P m_{G_i}$ columns in $\mathbf{B} \in \mathbb{R}^{L \times Q}$.

Let us note that obviously, the quality of the unmixing is going to be very dependent on the quality of the bundle itself. In some cases, a spectral library of endmembers incorporating some endmember variability may be available directly. In such a case, the methods we describe next still apply, and the unmixing paradigm becomes semi-supervised, as in the sparse unmixing approaches [24].

B. Abundance Estimation

Now that the bundles are available, one has to estimate the abundances of the different materials, which can be done in multiple ways. The Fisher Discriminant Nullspace (FDN) approach [34] is a dimension reduction technique based on finding a linear transformation which maps the data to a subspace where the intra class variability of the bundles is minimized, while their inter class variability is maximized. The unmixing is then carried out in this new subspace, e.g. using the centroids of each class as the endmembers.

Multiple Endmember Spectral Mixture Analysis (MESMA) [35] is a technique which aims at selecting in each pixel of the HSI the best endmember candidate for each material in terms of data fit. To do that, Problem (3) is solved using all possible combinations of (a small number of) candidate endmembers as the endmember matrix. However, the problem thus becomes combinatorial and is computationally expensive.

Machine learning-inspired approaches have also been designed, using the fact that “training data” can be constructed by simulating mixtures of the different instances of the materials in various proportions [36], [37].

Last, but not least, one can simply replace the endmember matrix \mathbf{S} with the bundle matrix \mathbf{B} in Eq. (3) and solve for the new abundance matrix in the exact same way, via constrained least squares. This approach is the most straightforward, but differs from the ones mentioned above insofar as several endmembers from each bundle can be simultaneously active, and summing the individual contributions of each material

¹This denomination should not be understood in the sense of the algebraic structure. Instead, this simply means that the endmembers (and hence the abundance coefficients) are grouped into P clusters, which we call groups here.

in each pixel is required to recover the abundances of the materials. This can be seen as a drawback, which makes the interpretation of the results harder. Of course, when the dictionary is not extracted from the data, and each instance represents a particular physical configuration of a material, this could indeed be an issue if the goal is to associate each pixel with a given physical parameter (though it is technically possible that several configurations of the same material can be present at the same time in a given pixel). Here, we show that this solution is actually theoretically more powerful than considering one instance per class, because it allows to define pixelwise endmembers which can be derived from the representatives of each material, and interpreted geometrically.

Indeed, we can write the LMM in the context of FCLSU with spectral bundles in one pixel (we drop the pixel index here to keep the notation uncluttered) in two different ways:

$$\mathbf{x} = \sum_{m=1}^Q a_m \mathbf{b}_m = \sum_{p=1}^P \left(\sum_{i=1}^{m_{G_p}} a_{G_p,i} \mathbf{b}_{G_p,i} \right), \quad (4)$$

where \mathbf{b}_m is the m^{th} column of this dictionary, and $a_{G_p,i}$ is the abundance coefficient associated to $\mathbf{b}_{G_p,i}$. Now, if we want the global abundance of material i to be $a_p = \sum_{i=1}^{m_{G_p}} a_{G_p,i}$, then we have to rewrite Eq. (4) as:

$$\mathbf{x} = \sum_{p=1}^P \left(\sum_{i=1}^{m_{G_p}} a_{G_p,i} \right) \left(\frac{\sum_{i=1}^{m_{G_p}} a_{G_p,i} \mathbf{b}_{G_p,i}}{\sum_{i=1}^{m_{G_p}} a_{G_p,i}} \right) = \sum_{p=1}^P a_p \mathbf{s}_p, \quad (5)$$

with $a_p \triangleq \sum_{i=1}^{m_{G_p}} a_{G_p,i}$ the total abundance coefficient for material p in the considered pixel, and $\mathbf{s}_p \triangleq \frac{\sum_{i=1}^{m_{G_p}} a_{G_p,i} \mathbf{b}_{G_p,i}}{\sum_{i=1}^{m_{G_p}} a_{G_p,i}}$. This vector actually contains a new ‘‘equivalent’’ endmember for this pixel and material p , associated with the global ‘‘intuitive’’ abundance coefficients. As a matter of fact, this new endmember is a weighted mean of all the available instances of this material, where the weights are the abundances extracted by FCLSU. Therefore, the normalized coefficients of this weighted mean are nonnegative and sum to one. This means that \mathbf{s}_p is a convex combination of the instances of the corresponding endmember. Geometrically, each equivalent endmember belongs to the convex hull of the elements of the bundle for this material. In this sense, finding abundances with FCLSU rather than MESMA allows more freedom in terms of spectral variability: the latter constrains each pixel to come from a combination of the extracted sources, while the former theoretically allows any point inside the convex hull of the each bundle to be a local endmember. This geometric interpretation is shown in Fig. 1. The per-pixel equivalent endmember \mathbf{s}_p is of course only defined if at least one instance of group G_p is active in this pixel. Otherwise, it makes no sense trying to extract spectral variability in a pixel from a material which is not present.

Similarly to what is done in the semi-supervised sparse unmixing paradigm, it makes sense to assume that only a few atoms of the bundle dictionary are going to be active in each pixel. Hence, one can add a regularization term to the

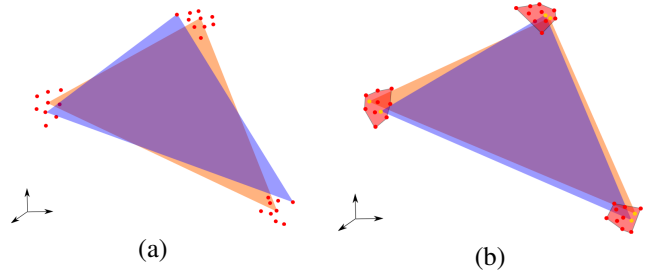


Fig. 1. Geometric interpretations of spectral unmixing with bundles. Two simplices are represented, each explaining a different pixel. (a) How endmember bundles are classically represented and (b) geometric interpretation of using FCLSU on the whole extracted dictionary. The red polytopes are the convex hull of the different bundles. The yellow points are accessible endmembers when using FCLSU, whereas they were not extracted by the EEA.

formulation (3) so as to enforce sparse abundance vectors:

$$\arg \min_{\mathbf{A} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{B}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1 \quad (6)$$

where $\|\cdot\|_1$ denotes the sparsity-inducing \mathcal{L}_1 norm of the matrix, i.e. the sum of the absolute values of all its entries, and λ is a regularization term, weighting the sparsity related term w.r.t. the data fit term. Note that the ASC has been dropped here; we explain below why.

This approach is useful to induce sparsity in the estimated abundances, but lacks two important features:

- The compatibility with the sum-to-one constraint. Indeed, let us assume that the ASC is enforced in problem (6). Since the abundances are constrained to be nonnegative, a straightforward computation shows that $\|\mathbf{A}\|_1 = N$. The \mathcal{L}_1 norm of the abundance matrix is then constant, and then minimizing this quantity will result in breaking the constraint. This phenomenon is known and has been reported for instance in [38], where the nonconvex $\mathcal{L}_{\frac{1}{2}}$ quasynorm instead is used to induce sparsity in the solutions.
- The structure of the bundles is not taken into account during the unmixing. Each abundance coefficient is treated independently w.r.t. sparsity. The first reason why sparsity was brought to hyperspectral image unmixing is the assumption that a few *materials*, rather than *atoms* of a dictionary, are active in each pixel.

III. PROPOSED ABUNDANCE ESTIMATION METHOD

A. Group sparsity inducing mixed norms

All the penalties we are going to detail in the following are based on applying a mixed norm on the abundance vector (in each pixel), which is endowed with a group structure G , which partitions the $Q = mP$ endmembers extracted by the n runs of VCA on random subsets into P groups (as many as the number of materials to unmix). We drop the pixel index for simplicity of the notation. In the most general form, the group two-level mixed $\mathcal{L}_{G,p,q}$ norm is defined, for any two positive

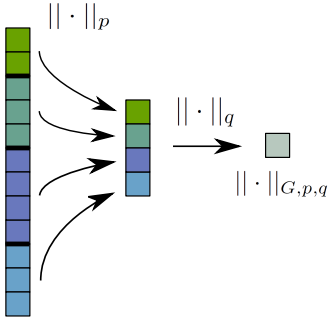


Fig. 2. Illustration on how the $\mathcal{L}_{G,p,q}$ norm operates on a vector, given the group structure G .

real numbers p and q as [29]:

$$\|\mathbf{a}\|_{G,p,q} \triangleq \left(\sum_{i=1}^P \left(\sum_{j=1}^{m_{G_i}} |a_{G_i,j}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} = \left(\sum_{i=1}^P \|\mathbf{a}_{G_i}\|_p^q \right)^{\frac{1}{q}}, \quad (7)$$

where \mathbf{a}_{G_i} is a subvector of the abundance vector (\mathbf{a}) comprising all the abundance coefficients associated to the endmembers of group G_i . This equation only defines a true norm for $p, q \geq 1$ (and also for p or $q = \infty$, by taking limits). As explained in Fig. 2, the idea is to take the p norm of each of the P subvectors of coefficient defined by the groups, to store the results in a P -dimensional vector, for which we are going to compute the q norm. We will see that with smart choices of p and q , different types of sparsity can be obtained when this mixed norm is used as a regularizer in the unmixing with bundles. The definition of the $\mathcal{L}_{G,p,q}$ norm can easily be extended to a matrix $\mathbf{A} \in \mathbb{R}^{Q \times N}$, using the same expression, operating columnwise, and summing the results on all pixels:

$$\|\mathbf{A}\|_{G,p,q} \triangleq \sum_{k=1}^N \|\mathbf{a}_k\|_{G,p,q}. \quad (8)$$

Here, we are interested in norms which can handle any group structure, while enforcing several types of sparsity. For instance, the use of sparsity in SU is based on the assumption that a few materials are active in each pixel. If the dictionary of endmembers has a group structure, it makes sense to enforce sparsity on the number of groups, rather than on the total number of signatures. This rationale is the basis of the so-called group LASSO [25]. This method uses the $\mathcal{L}_{G,2,1}$ norm, which enforces sparsity on the vector whose entries are the $\|\mathbf{a}_{G_i}\|_2$. This means that when one of these entries is zero, the whole group is discarded entirely since the vector \mathbf{a}_{G_i} has a zero norm. Within each group, there is no sparsity and thus most or all signatures are likely to be active. The effect of this penalty on a matrix $\mathbf{A} \in \mathbb{R}^{Q \times N}$ is shown in Fig. 3. In some cases, for example when we deal with a small number of groups, and we have reasons to believe that there is only one or few instances of each group which are active in a pixel, we may expect within-group sparsity, without group sparsity. In this case the elitist LASSO penalty [28] is suited to the problem, since it uses the $\mathcal{L}_{G,1,2}$ norm, which promotes a small value of the \mathcal{L}_1 norms of each \mathbf{a}_{G_i} . The effect of this penalty

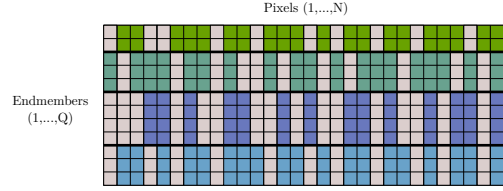


Fig. 3. Effect of the group LASSO penalty on the abundance matrix. The group structure is shown in colors (the rows of the matrix have been sorted for more clarity). Inactive entries of the matrix are in gray. A small number of groups is selected in each pixel, but within each group the matrix is dense.

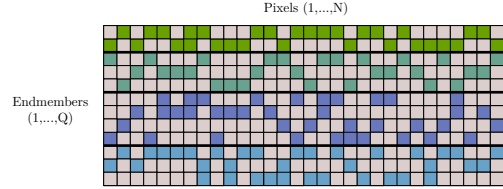


Fig. 4. Effect of the elitist LASSO penalty on the abundance matrix. A small number of instance in each group is selected in each pixel, but all or almost all groups are active.

is shown in Fig. 4. Using a penalty which enforces both group sparsity and global sparsity (on the total number of active signatures) also seems appealing. A seemingly natural way to do so with the proposed framework is to take both $p = q = 1$, to enforce sparsity on both norms that are computed in sequence. However, it turns out that the $\mathcal{L}_{1,1}$ norm actually is the same as the regular \mathcal{L}_1 norm, i.e. using these values for p and q actually breaks the group structure of the coefficients, and is of course still incompatible with the ASC. The sparse group LASSO [39] uses a combination of the group lasso penalty and a classical \mathcal{L}_1 norm to benefit from both properties. It was recently used in a sparse SU context in [26]. However, in this case, benefiting from both penalties comes at the cost of having two regularization parameters to tune. In our case, recall that this penalty is also still at odds with the ASC due to the presence of the \mathcal{L}_1 norm in the objective. The ASC can also be contradictory with sparsity in some other configurations: for instance, if each material has only one representative, the group LASSO reduces to the regular LASSO and the ASC conflicts with the objective. In order to avoid this issue, we are also using a fractional case, with the $\mathcal{L}_{G,1,q}$ “norm”, with $q = \frac{a}{b}$ (a and $b \in \mathbb{N}$) and $0 < \frac{a}{b} < 1$. This penalty is no longer a norm, because we lose convexity due to the fact that $q < 1$, but it has the advantage of enforcing both group sparsity and within-group sparsity in a compact formulation, without conflicting with the ASC anymore. In addition, the \mathcal{L}_q quasinorm $q < 1$ is a better approximation of the \mathcal{L}_0 norm than the \mathcal{L}_1 norm. The effect of the $\mathcal{L}_{G,1,q}$ penalty on the abundance matrix is shown in Fig. 5. The choice of the fraction is an additional parameter, but we will see that it turns out to be relatively easy to tune.

Geometrically, the group lasso encourages a small number of materials (i.e. groups) to be active in each pixel, but

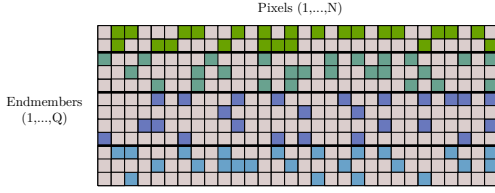


Fig. 5. Effect of the fractional LASSO penalty on the abundance matrix. A small number of instance in each group is selected in each pixel, and the mixture is also sparse within each group.

tends to prefer a dense mixture within each group. Hence, the pixelwise endmembers tend to lie around the center of mass of each bundle. For the elitist penalty, a large number of groups are active in each pixel. However, within each bundle, a small number of endmember candidates will be active, which translates geometrically into the fact that the pixelwise endmembers will lie inside polytopes with a few number of vertices. The fractional penalty has the same effect as the elitist one, except than in addition only a few groups are going to be active in each pixel.

B. Optimization

With either of those penalties, the optimization problem to solve is:

$$\arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{A}\|_{G,p,q} + \mathcal{I}_{\Delta_Q}(\mathbf{A}). \quad (9)$$

where, \mathcal{I}_{Δ_Q} is the indicator function of the probability simplex Δ_Q with Q vertices, applied independently to every column of \mathbf{A} . Note that after solving this problem, in order to obtain the global abundances, one simply has to sum the abundances within each instance of each group, as described in section II. The optimization problem (9) is convex for both the group and elitist penalties, but not for the fractional one.

In any case, since the problem is not differentiable, we are going to use the Alternating Direction Method of Multipliers (ADMM) [40] to solve it. Convergence to the global minimum is automatically guaranteed for the group and elitist penalties. For the nonconvex case, as we will see, the situation is more complex. The ADMM was designed to tackle convex problems, but it has been more and more (successfully) used for nonconvex problems as well, and recent works [41] show that if the nonconvex function satisfies some conditions, the ADMM is proven to converge to a stationary point in the nonconvex case. One of these cases includes the \mathcal{L}_p quasinorm for $p < 1$. This results remains to be shown in the mixed $\mathcal{L}_{G,1,q}$ norm with $q < 1$ case, (but it is likely to satisfy the same conditions, being “less nonconvex” than the \mathcal{L}_p quasinorm, because the unit ball of such a norm will have some nonconvex facets, but not all since some of them will be similar as the facets of the \mathcal{L}_1 ball).

The next two sections lay out the algorithm to solve the optimization problem with the group and elitist penalties, while the last one shows how to handle the fractional case.

1) *Group and elitist cases:* The optimization problem to solve in both cases is convex, and can be handled in the same fashion. In order to use the ADMM, we need to introduce auxiliary split variables, which will decompose problem (9) into easier subproblems. Let us then rewrite this problem as:

$$\begin{aligned} \arg \min_{\mathbf{A}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{U}\|_{G,p,q} + \mathcal{I}_{\Delta_Q}(\mathbf{V}) \\ \text{s.t.} \quad & \mathbf{U} = \mathbf{A}, \quad \mathbf{V} = \mathbf{A}. \end{aligned} \quad (10)$$

The ADMM consists in expressing the constrained problem defined in Eq. (10) in an unconstrained way using an Augmented Lagrangian (AL), and then minimizing it iteratively and alternatively for each of the variables introduced, before finally maximizing it w.r.t. the Lagrange Multipliers appearing in the AL (the so-called *dual update*). ρ is the *barrier* parameter weighting the AL terms (which we will set to $\rho = 10$ for all the algorithms). Here, the Augmented Lagrangian writes:

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \mathbf{U}, \mathbf{V}) = & \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{U}\|_{G,p,q} + \mathcal{I}_{\Delta_Q}(\mathbf{V}) \\ & + \frac{\rho}{2} \|\mathbf{A} - \mathbf{U} - \mathbf{C}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - \mathbf{V} - \mathbf{D}\|_F^2 \\ & - \frac{\rho}{2} \|\mathbf{C}\|_F^2 - \frac{\rho}{2} \|\mathbf{D}\|_F^2, \end{aligned} \quad (11)$$

where \mathbf{C} and \mathbf{D} are the set of dual variables. Each of the updates is supposed to be straightforward. The update w.r.t. \mathbf{A} only involves quadratic terms. and the updates w.r.t. \mathbf{U} and \mathbf{V} can be readily obtained, provided the proximal operators [42] of the mixed norm, and of the indicator function of the simplex can be easily computed. The latter can indeed be easily obtained with the algorithm of [43]. For both the group and elitist penalties, the proximal operators have closed forms, which we provide now.

The proximal operator of the group penalty is simply a group version of the so-called *block* soft thresholding operator, i.e. the proximal operator of the \mathcal{L}_2 norm:

$$\mathbf{prox}_{\tau \|\cdot\|_{G,2,1}}(\mathbf{v}) = \begin{bmatrix} \mathbf{soft}_{\tau}(\mathbf{v}_{G_1}) \\ \vdots \\ \mathbf{soft}_{\tau}(\mathbf{v}_{G_P}) \end{bmatrix}. \quad (12)$$

The block soft thresholding is denoted by \mathbf{soft}_{τ} , where τ is the scale parameter of this operator:

$$\forall \mathbf{u} \in \mathbb{R}^P, \mathbf{soft}_{\tau}(\mathbf{u}) = \left(1 - \frac{\tau}{\|\mathbf{u}\|_2}\right)_+ \mathbf{u}, \quad (13)$$

where $\cdot_+ = \max(\cdot, 0)$ (and we have $\mathbf{soft}_{\tau}(\mathbf{0}) = \mathbf{0}$).

The proximal operator for the elitist norm is a bit more complex, but has a closed form (derived in [44]), which involves the *regular* soft thresholding operator, the proximal operator of the \mathcal{L}_1 norm. We recall that

$$\mathbf{soft}_{\tau}(\mathbf{u})_i = \text{sign}(u_i)(|u_i| - \tau)_+. \quad (14)$$

With this definition, the proximal operator of the elitist penalty is given as:

$$\mathbf{prox}_{\tau \|\cdot\|_{G,1,2}}(\mathbf{v}) = \begin{bmatrix} \mathbf{soft}_{\gamma_1}(\mathbf{v}_{G_1}) \\ \vdots \\ \mathbf{soft}_{\gamma_P}(\mathbf{v}_{G_P}) \end{bmatrix}, \quad (15)$$

where the soft thresholding is applied entrywise, and

$$\gamma_i = \frac{\tau}{1 + \tau} \|\mathbf{v}_{G_i}\|_1, \quad \forall i \in \llbracket 1, P \rrbracket. \quad (16)$$

The ADMM procedure for any case is summarized in Algorithm 1, in which **prox** denotes the proximal operator of the function in index, and **proj** $_{\Delta_Q}$ denotes the proximal operator of the indicator function of the simplex, a simple projection. **I** denotes the identity matrix whose size is in index (we only provide one dimension for brevity since the matrix is square). Note that the proximal operators are carried out columnwise.

Data: \mathbf{X}, \mathbf{B}

Result: \mathbf{A}

Initialize \mathbf{A} and choose λ ;

while ADMM termination criterion is not satisfied **do**

$$\begin{aligned} \mathbf{A} &\leftarrow (\mathbf{B}^\top \mathbf{B} + 2\rho \mathbf{I}_Q)^{-1} (\mathbf{B}^\top \mathbf{X} + \rho(\mathbf{U} + \mathbf{V} + \mathbf{C} + \mathbf{D})); \\ \mathbf{U} &\leftarrow \mathbf{prox}_{(\lambda/\rho)\|\cdot\|_{G,p,q}}(\mathbf{A} - \mathbf{C}); \\ \mathbf{V} &\leftarrow \mathbf{prox}_{\mathcal{I}_{\Delta_Q}}(\mathbf{A} - \mathbf{D}) = \mathbf{proj}_{\Delta_Q}(\mathbf{A} - \mathbf{D}); \\ \mathbf{C} &\leftarrow \mathbf{C} + \mathbf{U} - \mathbf{A}; \\ \mathbf{D} &\leftarrow \mathbf{D} + \mathbf{V} - \mathbf{A}; \end{aligned}$$

end

Algorithm 1: ADMM to solve problem (10) in the convex cases).

2) *Fractional case:* The problem is more complex for the fractional mixed norm. As we have pointed out above, there is no proof that the mixed $\mathcal{L}_{G,1,q}$ norm with $q < 1$ satisfies the required properties for the ADMM to converge to a local minimum. However, in our problem, with an appropriate variable splitting scheme, we can express the fractional case for problem (9) as a \mathcal{L}_q regularized constrained least squares problem.

Let us suppose for simplicity (and without loss of generality), that the rows of \mathbf{A} and the columns of \mathbf{B} have been sorted so that, in each pixel, the abundance vector has the following form:

$$\mathbf{a} = [a_{1,1}, a_{1,2}, \dots, a_{1,m_{G_1}}, a_{2,1}, a_{2,2}, \dots, a_{2,m_{G_2}}, \dots, a_{G_P,1}, a_{G_P,2}, \dots, a_{G_P,m_{G_P}}]^\top.$$

Recall that m_{G_i} is the number of instances of one of the P groups, in this case G_i .

The problem we want to solve is:

$$\arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{A}\|_{G,1,q}^q + \mathcal{I}_{\Delta_Q}(\mathbf{A}), \quad (17)$$

with $Q = \sum_{i=1}^P m_{G_i}$ the total number of signatures in the bundle $\mathbf{B} \in \mathbb{R}^{L \times Q}$. With the following variable splitting scheme, Problem (17) is equivalent to:

$$\begin{aligned} \arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda \|\mathbf{U}\|_q^q + \mathcal{I}_{\Delta_Q}(\mathbf{V}) \\ \text{s.t. } \mathbf{MA} = \mathbf{U} \in \mathbb{R}^{P \times N}, \quad \mathbf{A} = \mathbf{V} \end{aligned} \quad (18)$$

with

$$\mathbf{M} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix}, \quad (19)$$

where $\mathbf{M} \in \mathbb{R}^{P \times Q}$, and the i^{th} row having m_{G_i} consecutive ones.

This way, we have reduced the optimization problem to a \mathcal{L}_q regularized least squares problem, where the variable on which the fractional norm is applied is a vector whose entries are the \mathcal{L}_1 norms of the abundance coefficients in each group. Note that the new problem is equivalent to the original one only thanks to the nonnegativity constraint, which allowed us to turn the \mathcal{L}_1 norm into linear constraints.

This way, the convergence of the ADMM for our nonconvex problem is, in theory, guaranteed [41], should we be able to compute exact updates for all the subproblems of the ADMM. However, even after this simplification of the problem, an issue remains: there is no closed form expression or known algorithm (to the best of our knowledge) to compute exactly the shrinkage operator of the \mathcal{L}_q quasinorm (to the power q) when $q < 1$, except when $q = \frac{1}{2}$ or $q = \frac{2}{3}$ [45]. Here, we prefer the term ‘‘shrinkage operator’’ to the term ‘‘proximal operator’’ because the latter is usually defined for convex functions only. In addition, this operator is a discontinuous function, because of the nonconvexity of the quasinorm [46]. This limits the applicability of proximal methods to solve this type of problems, and other types of algorithms (or of nonconvex sparsity inducing penalties) have been investigated in remote sensing (see e.g. [47] and references therein).

In our case, in order to be able to apply ADMM nonetheless, we need an explicit shrinkage operator. We resort to an approximate q -shrinkage operator $S_{q,\tau}$, as defined in [48]:

$$\forall \mathbf{u} \in \mathbb{R}^P, \quad S_{q,\tau}(\mathbf{x})_i = \text{sign}(x_i) (|x_i| - \tau^{2-q} |x_i|^{q-1})_+. \quad (20)$$

This operator reduces to the soft thresholding operator when $q = 1$ and to the *hard thresholding* operator when $q = 0$. The hard thresholding operator is closely related to the shrinkage operator of the \mathcal{L}_0 norm [48]. In addition, it can be shown ([48], theorem II.4) that the operator of Eq. (20) is actually the exact shrinkage operator of a nonconvex function (which we will denote as f_q) with desirable properties: it is separable w.r.t. each entry of \mathbf{x} , with $f_q(\mathbf{x}) = \sum_{i=1}^P g_q(x_i)$, and the function g_q is even, continuous, strictly increasing and concave for $x_i > 0$, differentiable everywhere except in 0, and satisfies the triangle inequality. This function behaves in a way similar to the absolute value for small values of its argument, and more like the absolute value taken to the power q for larger arguments (see [48] for a graphical representation). For $q = 1$, this penalty function is the absolute value, but in general, unfortunately, there is no analytical expression for it. This result is interesting because we have an explicit shrinkage operator with nice properties to use with any

proximal algorithm, to the cost of having a regularizer without an explicit expression. Nevertheless, the convergence of the ADMM in the nonconvex case remains to be proven, since we replaced the \mathcal{L}_q norm with another nonconvex penalty, which should itself satisfy the required properties of [41] in order to guarantee convergence.

Finally, the optimization problem we solve is:

$$\begin{aligned} \arg \min_{\mathbf{A}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda f_q(\mathbf{U}) + \mathcal{I}_{\Delta_Q}(\mathbf{V}) \\ \text{s.t.} \quad & \mathbf{MA} = \mathbf{U}, \quad \mathbf{A} = \mathbf{V}. \end{aligned} \quad (21)$$

The AL writes:

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \mathbf{U}, \mathbf{V}) = & \frac{1}{2} \|\mathbf{X} - \mathbf{BA}\|_F^2 + \lambda f_q(\mathbf{U}) + \mathcal{I}_{\Delta_Q}(\mathbf{V}) \\ & + \frac{\rho}{2} \|\mathbf{MA} - \mathbf{U} - \mathbf{C}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - \mathbf{V} - \mathbf{D}\|_F^2 \\ & - \frac{\rho}{2} \|\mathbf{C}\|_F^2 - \frac{\rho}{2} \|\mathbf{D}\|_F^2, \end{aligned} \quad (22)$$

and we can use the ADMM algorithm to minimize it. The optimization procedure is described in Algorithm 2 (the approximate q -shrinkage is performed coordinate-wise).

Data: \mathbf{X}, \mathbf{B}

Result: \mathbf{A}

Initialize \mathbf{A} and choose λ ;

while *ADMM termination criterion is not satisfied* **do**

$\mathbf{A} \leftarrow (\mathbf{B}^\top \mathbf{B} + \rho \mathbf{M}^\top \mathbf{M} + \rho \mathbf{I}_Q)^{-1} (\mathbf{B}^\top \mathbf{X} + \rho \mathbf{M}^\top (\mathbf{U} + \mathbf{V}) + \rho (\mathbf{C} + \mathbf{D})) ;$
$\mathbf{U} \leftarrow S_{q, \lambda / \rho}(\mathbf{MA} - \mathbf{C}) ;$
$\mathbf{V} \leftarrow \text{prox}_{\mathcal{I}_{\Delta_P}}(\mathbf{A} - \mathbf{D}) = \text{proj}_{\Delta_P}(\mathbf{A}_{\mathcal{R}} - \mathbf{D}) ;$
$\mathbf{C} \leftarrow \mathbf{C} + \mathbf{U} - \mathbf{MA} ;$
$\mathbf{D} \leftarrow \mathbf{D} + \mathbf{V} - \mathbf{A} ;$

end

Algorithm 2: ADMM to solve problem (21).

C. Computational complexity

For the three proposed algorithms, the most costly operations involved are matrix products and solving linear systems. Hence the computational complexity of the algorithms for the group and elitist penalties is $\mathcal{O}(Q^2(L + N))$ (assuming $Q \ll N$). The complexity for FCLSU is the same but a much faster convergence is expected since the constraints are less complex and much less iterations will be required to reach convergence. Since the fractional penalty involves additional matrix products using matrix \mathbf{M} , then the complexity for the corresponding algorithm is slightly higher and is $\mathcal{O}(Q^2(L + N + P) + QPN)$. A significant increase in running time is then expected for the fractional penalty when the number of endmembers P gets large.

IV. RESULTS

A. Synthetic data

First, we test the three proposed penalties in a semi-supervised scenario, where the possible instances of each

endmember are known beforehand. This case is similar to the now well-known and widespread sparse unmixing approaches [24], [49]. The main difference with respect to regular sparse unmixing is that here the spectral library incorporates several variants of each endmember in order to account for the variability of the scene. The group structure G is then known a priori as well. The objective of this dataset is to compare the performance of the proposed penalties in terms of abundance estimation and variability retrieval, i.e. the core of the contribution of this paper, when the bundle extraction is assumed to have been carried out efficiently.

In order to show the interest of the proposed penalties, we designed a synthetic dataset corresponding to a challenging unmixing scenario, with 20 different materials present in the scene. We then selected randomly $P = 20$ signatures from the United States Geological Survey (USGS) spectral library [50], comprising 224 spectral bands in the visible and near infra-red parts of the electromagnetic spectrum. For each of those 20 signatures, spectral variants were generated by modifying the spectra using several effects: scaling variations, accounting for local changes in illumination conditions [21], a nonlinear quadratic (in the original signature) perturbation to mimick intrinsic variability effects, and i.i.d white Gaussian noise. 20 such variants were randomly generated for each material, providing a dictionary of $Q = 400$ spectral signatures, divided into $P = 20$ groups. In each pixel, only one signature is active for each endmember. 50×50 sparse (between 1 and 3 active materials in each pixel) abundance maps were computed using Gaussian Random Fields, so as to give them a spatial coherence (which is not used by the algorithm but is helpful to visualize the results).

We tested the performance of 4 algorithms on this dataset: the FCLSU algorithm, which does not take the group structure into account, and the three proposed penalties: group penalty, elitist penalty, and fractional penalty. In each case, we optimize the regularization parameter via a grid search to obtain the best abundance estimation performance ($RMSE(\hat{\mathbf{A}})$, see below). The FCLSU algorithm was implemented using ADMM as well, to allow a fair comparison for the running times of the various algorithms. We ran all the algorithms for 1000 iterations or stopped them when the relative variations (in norms) of the abundance matrix went below 10^{-6} . For the fractional penalty, the grid search is also carried out w.r.t. the choice of the fraction (we tested values between $\frac{1}{10}$ and $\frac{9}{10}$, with steps of $\frac{1}{10}$). To quantitatively assess the abundance estimation performance, we use the abundance Root Mean Squared Error $RMSE(\hat{\mathbf{A}})$ between the estimated abundances $\hat{\mathbf{a}}_k$ for each material (summing the abundances of all the representatives for each group) and the true ones \mathbf{a}_k in each pixel:

$$RMSE(\hat{\mathbf{A}}) = \frac{1}{N} \sum_{k=1}^N \sqrt{\frac{1}{P} \sum_{p=1}^P (a_{pk} - \hat{a}_{pk})^2} \quad (23)$$

where the abundance matrix comprises only the global abundance coefficients.

To assess the precision of the abundance estimation within each group, we also use this metric for each individual atom of the dictionary, which we call in that case $RMSE_G(\hat{\mathbf{A}})$:

$$RMSE_G(\hat{\mathbf{A}}) = \frac{1}{N} \sum_{k=1}^N \sqrt{\frac{1}{P} \sum_{p=1}^{P \times m_G} (a_{pk} - \hat{a}_{pk})^2}, \quad (24)$$

where $m_G = 20$ is the number of instances in each group, and here the abundance matrix used is the full bundle dictionary.

To compare the performance in terms of material variability retrieval, we also use the Root Mean Squared Error between the true signatures and the estimated ones:

$$RMSE(\hat{\mathbf{S}}) = \frac{1}{NP} \sum_{k=1}^N \sum_{p=1}^P \frac{1}{\sqrt{L}} \|s_{pk} - \hat{s}_{pk}\|_2 \quad (25)$$

and the overall spectral angle mapper (SAM) between the true signature \mathbf{b}_{pk} used for material p in pixel k and the estimated ones $\hat{\mathbf{s}}_{pk}$ (using Eq. (5)):

$$SAM = \frac{1}{NP} \sum_{k=1}^N \sum_{p=1}^P \arccos \left(\frac{\mathbf{b}_{pk}^\top \hat{\mathbf{s}}_{pk}}{\|\mathbf{b}_{pk}\|_2 \|\hat{\mathbf{s}}_{pk}\|_2} \right) \quad (26)$$

All the quantitative results are gathered in Table I.

From this table, we see that the elitist algorithm obtains the worst performance in terms of abundance estimation, followed by the FCLSU algorithm, while the group penalty largely outperforms those two when it comes to global abundances. The fractional penalty is able to refine these results slightly and obtains the best performance. If we look at what happens within groups, then the fractional penalty is outperformed by the others. We will see why it is so when we look at the abundance maps. The group and fractional penalties obtain the best endmember RMSE and SAM values, meaning that they recover the closest endmembers to the ground truth. In this configuration, the inter-group sparsity property of the abundances for those two penalties is crucial. The reason the fractional penalty further improves the estimation is that it does not encourage a too harsh within group sparsity, while enforcing inter group sparsity. On the other hand, this sparsity comes at the expense of a precise identification of which endmember candidates were actually used in each pixel. In terms of running time, except for FCLSU which is the fastest, the other approaches are notably slower. The fractional penalty is without surprise the slowest because the constraints in the optimization problem are more complex than for the other two penalties.

The optimal value of $\frac{a}{b}$, among those we tested, for the fractional penalty is $\frac{1}{10}$ possibly because this makes the sparsity inducing term closer to the group $\mathcal{L}_{1,0}$ norm. This shows the interest of the proposed approximation scheme to optimize this mixed fractional norm for any value of $\frac{a}{b}$, instead of simply settling with the cases where $\frac{a}{b} = \frac{1}{2}$ or $\frac{a}{b} = \frac{2}{3}$. The sensitivity to this parameter is not very critical, as can be seen in Fig. 6, where the optimal (over λ) $RMSE(\hat{\mathbf{A}})$

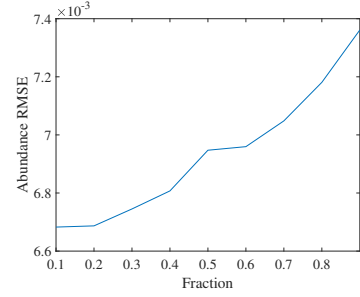


Fig. 6. Optimal value of $RMSE(\hat{\mathbf{A}})$ as a function of the fraction used for the fractional penalty.

is reported for the different fraction values.

Fig. 7 shows the abundance maps for 5 of the 20 materials to unmix, for all the tested algorithms. The abundances for the FCLSU algorithm and the group penalty look very similar, with a slight improvement in terms of “noisiness” of the abundance maps for the latter (e.g. for material 3). The abundances for the elitist penalty are slightly worse than the other methods, because although the elitist penalty encourages within group sparsity, it also tends to favor dense mixtures over the groups. The fractional penalty obtains the best visual results, by eliminating most of the noise in the abundances, and getting the support of each material right. However, it seems that it tends to locally make the abundances slightly too close to 1.

We also show in Fig. 8 a part of the abundance matrices obtained as images, to show the structure that each penalty induces on the coefficients. We confirm that FCLSU and the group penalty obtain similar performance, while the elitist penalty tends to incorporate spurious endmembers because of the between group density of the coefficients. The fractional penalty obtains a more accurate support of the endmembers than the other two, but at the price of sometimes overestimating the abundances because of a too violent sparse behavior, which is necessary to get rid of the noise but too harsh for a high precision abundance estimation.

B. Real data

In this section, we test the proposed algorithms on a real hyperspectral image. The dataset is a subset of an image acquired over the University of Houston campus, Texas, USA, in June 2012. It was used in the 2013 IEEE GRSS Data Fusion Contest (DFC) [51]. The image comprises 144 spectral bands in the 380 nm to 1050 nm region, and comes with a LiDAR dataset acquired a day before over the same area, with the same spatial resolution (2.5 m). We are interested here in a $152 \times 108 \times 144$ subset of this image, acquired over Robertson stadium on the Houston Campus and its surroundings. Fig. 9 shows a RGB representation of the observed scene, as well as a high spatial resolution RGB image of the scene, and a LiDAR-derived digital elevation model. This dataset was used in several works on spectral unmixing and has been shown to exhibit significant variability effects [21]. It comprises

Metric \ Algorithm	FCLSU	Group	Elitist	Fractional
$RMSE(\hat{\mathbf{A}})$	0.0103	0.0072	0.0158	0.0064
$RMSE_G(\hat{\mathbf{A}})$	0.0166	0.0183	0.0169	0.0209
$RMSE(\hat{\mathbf{S}})$	0.0342	0.0339	0.0382	0.0337
SAM (degrees)	2.026	1.749	2.148	1.738
Running time (s)	16	73	74	95
λ	\times	0.006	0.01	0.02

TABLE I

VALUES OF THE DIFFERENT METRICS FOR EACH TESTED ALGORITHM ON THE SYNTHETIC DATA. THE BEST VALUE IS RED, AND THE SECOND BEST IS IN BLUE. THE VALUES OF THE REGULARIZATION PARAMETERS ARE ALSO REPORTED , WHEN APPLICABLE.

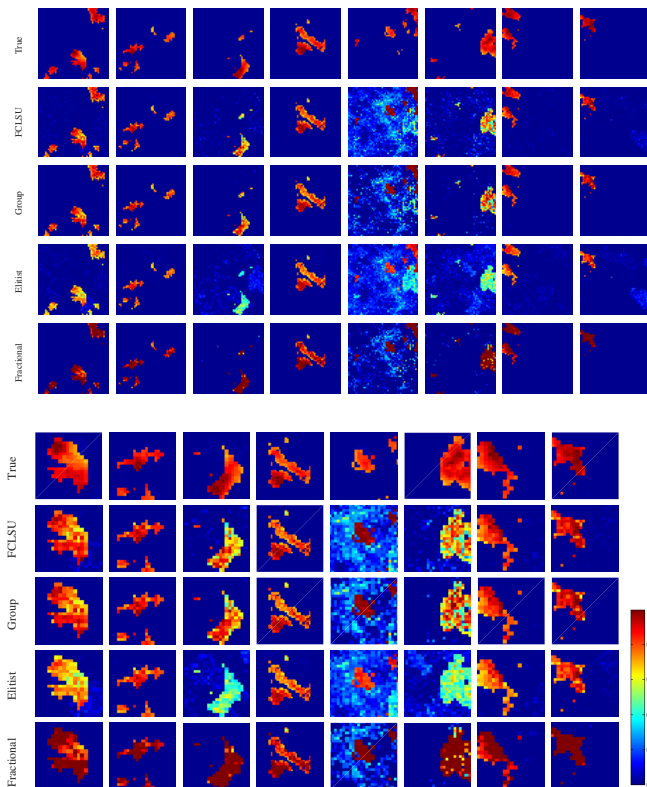


Fig. 7. Abundances for 8 of the 20 materials, for all the tested algorithms and the ground truth (top rows). The bottom figure is a zoomed version of the abundance maps to better highlight the behavior of the different algorithms.

several structures with locally changing geometries, namely the red metallic roofs which are pyramidal and whose facets correspond to different incidence and emergence angles from the light. The concrete stands also show different orientations with respect to the sun. Vegetation also incorporates variability, since it can be divided into two subclasses: trees and grass (which can in addition can be mixed with soil, which is interpreted as variability in the absence of a soil endmember).

Since a library of reference endmembers is unavailable for this image (let alone incorporating variability), the unmixing chain has to be completely blind and we resort to the AEB method to extract bundles. We sample without replacement 10 subsets comprising 10% of the data pixels each, and extract $P = 5$ endmembers on each subset using the VCA. Then we cluster the resulting dictionary into $P = 5$ classes using

the k-means algorithm (with the cosine similarity measure). Therefore, in this case, we have $Q = 50$. We then compare the performance of the different algorithms in several ways. We use $\frac{a}{b} = \frac{1}{10}$ for the fractional algorithm. The values of the regularization parameters are reported in Table II.

We first show the abundance maps obtained for each algorithm in Fig. 10. From those, we see that all algorithms seem to be able to reasonably explain the material variability present within the image, as opposed to using classical linear unmixing algorithms [21]. The group penalty is able to sparsify the abundance maps more than FCLSU does, providing clearer abundance maps for the red roofs and the painted structures which are next to the left and right stands of the stadium. However, the detection of the concrete stands themselves is slightly less accurate. The elitist penalty is able to identify the different structures in the image, but fails to consider them as pure because it favors a large number of groups to be simultaneously active in each pixel. The fractional penalty logically provides sparser abundance maps because it penalizes dense mixtures over the groups more aggressively than the group penalty. The structures are considered as purer, which makes sense for an urban scenario such as this one. However, the abundance maps are slightly noisier, with some neighboring pixels in homogeneous regions which do not share the same sparsity properties. This could be mitigated by using spatial regularizations on the abundance maps.

Then we show in more detail the abundance maps of 5 of the 10 representatives of the bundles corresponding to vegetation (Fig. 11), red roofs (Fig. 12), and concrete (Fig. 13). The idea is to find out if the various algorithms are able to provide a meaning to the different representatives within each bundle. First, for all materials, it seems that the group penalty tends to favor dense mixtures within each group, as expected, making it hard to give an interpretation to each endmember candidate. On the other hand, the elitist penalty obtains much sparser abundance maps within one group, and some endmember candidates can be interpreted (the football field is assigned to a single endmember candidate in the vegetation class, for example). The same goes for FCLSU, which seems to naturally enforce a certain sparsity within each group. Finally, the fractional penalty provides the most conclusive results. For each class, only the predominant signatures are retained and are interpretable, whereas the others are very sparse and some are almost entirely discarded (because they

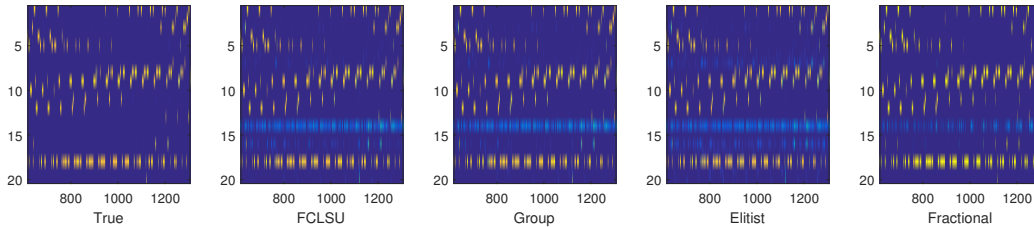


Fig. 8. Abundance matrices (pixels are on the x-axis, and candidate endmembers are on the y-axis. Candidates 1-5 belong to group 1, candidates 5-10 to group 2, and so on).

Metric \ Algorithm	Algorithm			
	FCLSU	Group	Elitist	Fractional
RMSE	0.0029	0.010	0.016	0.008
SAM (degrees)	0.7240	2.7891	5.1944	2.3684
Running time (s)	3	8	51	63
λ	\times	2	0.4	0.5

TABLE II

VALUES OF THE RMSE AND SAM (COMPUTED ON THE RECONSTRUCTION) AND RUNNING TIMES FOR EACH TESTED ALGORITHM ON THE HOUSTON DATA. THE BEST VALUES ARE IN RED, AND THE SECOND BEST ARE IN BLUE. THE VALUES OF THE REGULARIZATION PARAMETERS ARE ALSO REPORTED, WHEN APPLICABLE.

could be spurious, or redundant). For the vegetation class, three subclasses are identified by the algorithm (only two are clearly identifiable for FCLSU): football field, trees, and the area within the field which is mixed with soil. This is interpreted as variability by the model. The other less explicative abundance maps are made more sparse by the algorithm. For the red roof class, the facets of each roof are assigned to two different categories, depending on their orientation w.r.t. the sun, and a third map detects the other red roofs of the image, which are subject to other illumination conditions. The concrete stands are assigned to two different signatures for similar reasons. Note that the abundances of these subclasses are rarely exactly one; this means that the algorithms prefers to identify as a local endmember a convex combination of these two signatures, thus taking into account variability effects which are not sufficiently described by the endmember candidates only.

Finally, for completeness, we show in Table II the overall RMSE and SAM values for each algorithm (computed on the reconstruction of the pixels, in the absence of ground truth), the running times as well as the chosen values for the regularization parameter, when applicable. Without surprise, the RMSE and SAM values decrease when either of the three proposed penalties are used, since we compromise between accurate reconstruction of the data and structured sparsity of the abundance maps. We still note that apart from FCLSU the reconstruction errors are the best when the fractional penalty is used. Since P is rather small here, the running times are more comparable for the different algorithms, except that FCLSU and the group penalty converge after fewer iterations.

V. CONCLUSION

In this paper, we have proposed and compared three different sparsity-inducing penalties, specifically aimed at tackling endmember variability in hyperspectral image unmixing, when

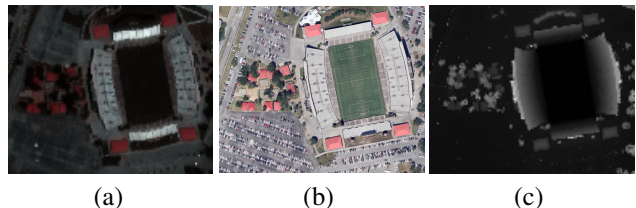


Fig. 9. A RGB representation of the Houston hyperspectral dataset (a). High spatial resolution color image acquired over the same area at a different time (b). Associated LiDAR data (c), where black corresponds to 9.6m and white corresponds to 46.2m.

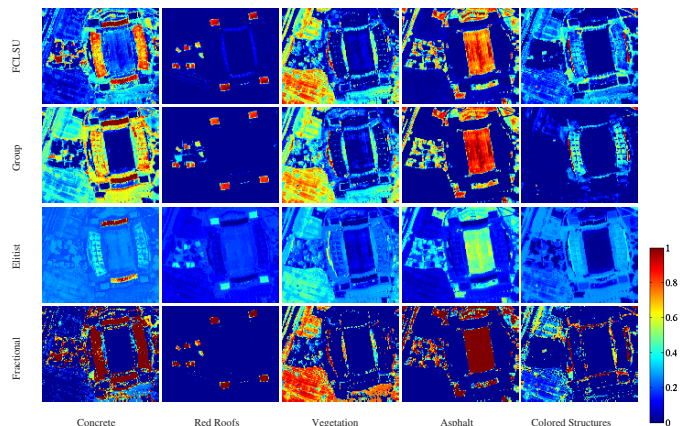


Fig. 10. Abundances for all the tested algorithms on the Houston dataset.

a dictionary of endmember candidates is available, or when endmember bundles are extracted from the data. We provided a new geometric interpretation of unmixing hyperspectral data within this paradigm. Among the three proposed penalties, the group one enforces a natural inter-class sparsity property of the abundance maps, but is hampered by the fact that it prefers dense mixtures within each group. Conversely, the elitist

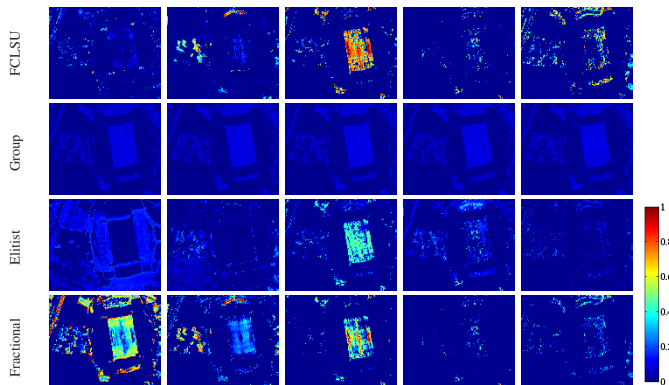


Fig. 11. Abundances for a few vegetation endmember candidates, for all the tested algorithms.

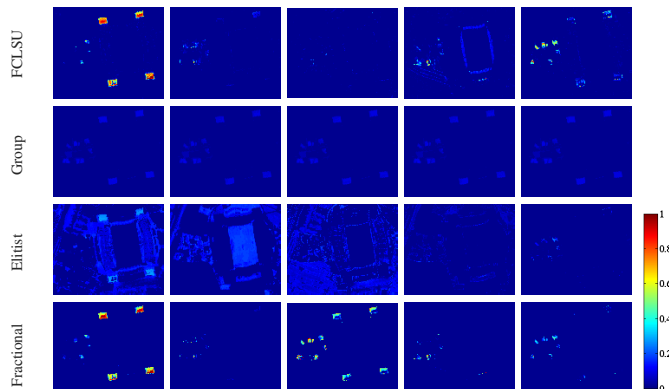


Fig. 12. Abundances for a few red roofs endmember candidates, for all the tested algorithms.

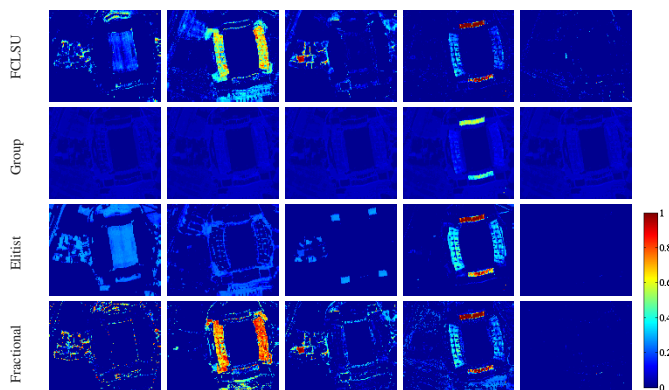


Fig. 13. Abundances for a few concrete endmember candidates, for all the tested algorithms.

penalty tries to select the best endmember candidates within each group, but provide inconclusive results because it prefers dense mixtures over the groups. The new fractional penalty we propose leads to the best results: it allows to enforce both within and inter group sparsity at the same time in a single mixed quasinorm, and is compatible with the sum-to-one constraint on the abundances. The interest of allowing to use any fraction is to provide more flexibility and better results than using the two cases the shrinkage operators have a closed form ($\frac{a}{b} = \frac{1}{2}$, or $\frac{a}{b} = \frac{2}{3}$). It finally provides purer global abundance maps while giving clearer interpretations to subclasses than the other tested methods. Future work could include the combination of this penalty to spatial regularizations on the abundances, as well as an automatic way to tune the regularization parameter providing a compromise between sparsity and data fit. Finding a proof of convergence of the Alternating Direction Method of Multipliers (ADMM) in the approximation framework we use for the optimization would also be an interesting perspective.

REFERENCES

- [1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, pp. 6–36, June 2013.
- [2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44–57, Jan 2002.
- [3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, pp. 354–379, April 2012.
- [4] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, pp. 67–81, Jan 2014.
- [5] A. Robin, K. Cawse-Nicholson, A. Mahmood, and M. Sears, "Estimation of the intrinsic dimension of hyperspectral images: Comparison of current methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, pp. 2854–2861, June 2015.
- [6] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 650–663, March 2004.
- [7] J. Nascimento and J. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 898–910, April 2005.
- [8] D. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 529–545, Mar 2001.
- [9] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 1844–1868, June 2014.
- [10] N. Dobigeon, J. Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, pp. 82–94, Jan 2014.
- [11] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, pp. 95–104, Jan 2014.
- [12] L. Drumetz, J. Chanussot, and C. Jutten, "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, 2016.
- [13] 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*, vol. 25, pp. 4565–4579, Oct 2016.

- [14] L. Drumetz, B. Ehsandoust, J. Chanussot, B. Rivet, M. Babaie-Zadeh, and C. Jutten, "Relationships between nonlinear and space-variant linear models in hyperspectral image unmixing," *IEEE Signal Processing Letters*, vol. 24, pp. 1567–1571, Oct 2017.
- [15] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1083–1094, 2000.
- [16] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, pp. 396–408, April 2012.
- [17] M. Xu, L. Zhang, and B. Du, "An image-based endmember bundle extraction algorithm using both spatial and spectral information," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2607–2617, 2015.
- [18] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 525–538, 2016.
- [19] Y. Zhou, A. Rangarajan, and P. D. Gader, "A gaussian mixture model representation of endmember variability in hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2018.
- [20] A. Halimi, N. Dobigeon, and J. Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Transactions on Image Processing*, vol. 24, pp. 4904–4917, Dec 2015.
- [21] 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*, vol. 25, pp. 3890–3905, Aug 2016.
- [22] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, pp. 3219–3232, July 2016.
- [23] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Online unmixing of multitemporal hyperspectral images accounting for spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 3979–3990, 2016.
- [24] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 2014–2039, June 2011.
- [25] L. Meier, S. Van De Geer, and P. Bühlmann, "The group LASSO for logistic regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 53–71, 2008.
- [26] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral unmixing with sparse group LASSO," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pp. 3586–3589, IEEE, 2011.
- [27] T. R. Meyer, L. Drumetz, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral unmixing with material variability using social sparsity," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2187–2191, Sept 2016.
- [28] M. Kowalski and B. Torrèsani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," *Signal, image and video processing*, vol. 3, no. 3, pp. 251–264, 2009.
- [29] M. Kowalski, K. Siedenburg, and M. Dorfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2498–2511, 2013.
- [30] M. Goenaga, M. Torres-Madronero, M. Velez-Reyes, S. Van Bloem, and J. China, "Unmixing analysis of a time series of hyperion images over the guanica dry forest in puerto rico," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, pp. 329–338, April 2013.
- [31] K. Canham, A. Schlamm, A. Ziemann, B. Basener, and D. Messinger, "Spatially adaptive hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 4248–4262, Nov 2011.
- [32] M. A. Veganzones, G. Tochon, M. Dalla Mura, A. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation," *IEEE Transactions on Image Processing*, vol. 23, pp. 3574–3589, Aug 2014.
- [33] G. Tochon, L. Drumetz, M. A. Veganzones, M. Dalla Mura, and J. Chanussot, "From local to global unmixing of hyperspectral images to reveal spectral variability," in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2016)*, 2016.
- [34] J. Jin, B. Wang, and L. Zhang, "A novel approach based on fisher discriminant null space for decomposition of mixed pixels in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, pp. 699–703, Oct 2010.
- [35] D. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 267 – 279, 1998.
- [36] F. A. Mianji and Y. Zhang, "Svm-based unmixing-to-classification conversion for hyperspectral abundance quantification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 4318–4327, Nov 2011.
- [37] T. Uezato, R. J. Murphy, A. Melkumyan, and A. Chlingaryan, "Incorporating spatial information and endmember variability into unmixing analyses to improve abundance estimates," *IEEE Transactions on Image Processing*, vol. 25, pp. 5563–5575, Dec 2016.
- [38] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 6793–6806, Nov 2014.
- [39] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group LASSO," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.
- [40] 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.
- [41] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *arXiv preprint arXiv:1511.06324*, 2015.
- [42] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212, Springer, 2011.
- [43] L. Condat, "Fast projection onto the simplex and the \mathcal{L}_1 ball," *Mathematical Programming*, pp. 1–11, 2014.
- [44] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.
- [45] W. Cao, J. Sun, and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of regularization," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 31–41, 2013.
- [46] M. Yukawa and S. I. Amari, " \mathcal{L}_p -regularized least squares ($0 < p < 1$) and critical path," *IEEE Transactions on Information Theory*, vol. 62, pp. 488–502, Jan 2016.
- [47] D. Tuia, R. Flamary, and M. Barlaud, "Nonconvex regularization in remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6470–6480, 2016.
- [48] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, p. 075004, 2016.
- [49] Y. Altmann, M. Pereyra, and J. Bioucas-Dias, "Collaborative sparse regression using spatially correlated supports - application to hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 24, pp. 5800–5811, Dec 2015.
- [50] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. M. Benzil, H. A. Lowers, R. L. Driscoll, et al., "Usgs spectral library version 7," tech. rep., US Geological Survey, 2017.
- [51] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 2405–2418, June 2014.