

# Inertia-Constrained Pixel-by-Pixel Nonnegative Matrix Factorisation: a Hyperspectral Unmixing Method Dealing with Intra-class Variability

Charlotte Revel, Yannick Deville, *Member, IEEE*, Véronique Achard and Xavier Briottet

**Abstract**—Blind source separation is a common processing tool to analyse the constitution of pixels of hyperspectral images. Such methods usually suppose that pure pixel spectra (endmembers) are the same in all the image for each class of materials. In the framework of remote sensing, such an assumption is no more valid in the presence of intra-class variabilities due to illumination conditions, weathering, slight variations of the pure materials, etc... In this paper, we first describe the results of investigations highlighting intra-class variability measured in real images. Considering these results, a new formulation of the linear mixing model is presented leading to two new methods. Unconstrained Pixel-by-pixel NMF (UP-NMF) is a new blind source separation method based on the assumption of a linear mixing model, which can deal with intra-class variability. To overcome UP-NMF limitations an extended method is proposed, named Inertia-constrained Pixel-by-pixel NMF (IP-NMF). For each sensed spectrum, these extended versions of NMF extract a corresponding set of source spectra. A constraint is set to limit the spreading of each source's estimates in IP-NMF. The methods are tested on a semi-synthetic data set built with spectra extracted from a real hyperspectral image and then numerically mixed. We thus demonstrate the interest of our methods for realistic source variabilities. Finally, IP-NMF is tested on a real data set and it is shown to yield better performance than state of the art methods.

**Index Terms**—Nonnegative Matrix Factorisation (NMF), Blind source separation, Hyperspectral unmixing, intra-class variability.

## I. INTRODUCTION

**H**YPERSPECTRAL imaging is a common tool in the framework of remote sensing. Images provided by these sensors are spectrally highly resolved. However they lead to a decrease of the spatial resolution. Thus each signal recorded by a pixel is the result of a combination of spectra of several pure materials composing the area delimited by the pixel projected onto the ground. Retrieving these pure spectra, usually called endmembers, and each material proportion, called abundances, brings, in each considered pixel, interesting sub-pixel information. This source separation problem is called unmixing in remote sensing. How pure reflectance spectra are combined depends on the scene. The most common mixing model is the linear one [1]. The weight of each pure material

spectrum in each sensed signal is proportional to the covered area. This model is adapted to flat macroscopic observed scenes [2] under a homogeneous irradiance. In case of 3D scene, nonlinear models are often more adapted, for instance a linear-quadratic mixing model was developed in [3] for urban areas. For microscopic effects, intimate mixture models are used [4]. Once the nature of the mixing model is known, source separation, also named unmixing in this study, can be performed. A large number of methods, requiring more or less prior knowledge, have been developed [5]. However they hardly handle source variations. Source variations mean that the reflectance spectrum of a source of a same class can vary from an observation to another. For instance, the spectra of two similar roof materials in two different locations may differ. In remote sensing, this phenomenon is called intra-class variability. Various approaches [6] have been developed to deal with this problem. Some of them make use of library [7]. Others described the intra-class variability as a statistical distribution and have proposed Bayesian methods, with various prior knowledges and models, to find the characteristics of this distribution and the associated abundances [8], [9]. However they frequently need particular knowledges. In a recent review of the unmixing methods dealing with intra-class variability [6], only two methods do not use prior knowledge or pre-defined source libraries. These two methods create a library from observed data and perform a sparse regression with this source library [10]. Also Somers et al., in [11], proposed an unmixing method based on a library extraction. This type of approach is interesting since a library allows one to describe source variability. Indeed if the library is large enough it can contain several spectra of the same material. However it does not allow one to extract exact sources in each pixel signal.

Here we propose a blind source separation method to extract from each observed signal the exact sources composing it. Our final method, called IP-NMF, is a constrained evolution of UP-NMF (Unconstrained Pixel-by-pixel NMF) that we first introduce. UP-NMF is based on an extended Nonnegative Matrix Factorisation (NMF) introduced by Lee and Seung in [12]. Section II evaluates the intra-class variability problems on a real image. Section III describes the general mixing model formulation ensuing from the observations developed in the previous part, Sections IV and V, our extended UP-NMF and IP-NMF methods, Section VI, their performance on a realistic data set and Section VII the performance of IP-NMF on a real image.

The research presented in this paper was funded by the Midi-Pyrénées region and the French ANR (ANR HYPE ANR 14-CE22-0016-01).

Ms. Revel and Mr. Deville are with the department of Signal, Image in Universe Sciences, IRAP, Toulouse University, 31400, France.

Ms. Revel, Ms. Achard and Mr. Briottet are with the department of Theoretical and Applied Optics, ONERA The French Aerospace Lab, Toulouse University, Toulouse, 31400, France.

## II. INTRA-CLASS VARIABILITY

### A. Problem statement

The popular unmixing model described by Keshava and Mustard in [1] assumes that a scene can be fully reconstructed by using one spectrum per pure material present in the image, called an endmember. The notion of material is hard to define when the information of interest to be extracted from the scene has a macroscopic scale. Remote sensing frequently aims at studying landscapes, in this case what is called an endmember, or source, is a class of materials rather than a chemical material. For instance endmembers may be: grass, roads, tiles, trees... Now the grass is not the same from an area to another (various species, hydric stress...), tiles vary from one roof to another (various weatherings, mineral slight composition variations ...), and so on [13], [14]. So, at a macroscopic scale, an endmember is not the spectrum of a pure material but the spectrum which better characterises a class of materials. Spectra depicted in Figure 1 corroborate this conclusion.

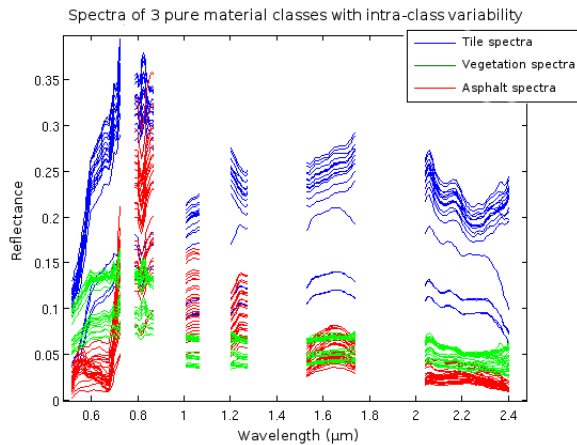


Fig. 1. Spectra of 3 different classes of materials: vegetation (red), tile (blue), asphalt (green).

Using this new definition of an endmember, intra-class variability is the spectral variability between reflected spectra present in the image and the endmember associated with them. Currently, intra-class variability is modelled in different ways. In [15] and [16] this variability is described by a scale factor. With this model the set  $\mathcal{E}_m$  of spectra, present in the image, belonging to the  $m^{\text{th}}$  source, is written as follows:  $\mathcal{E}_m = \{\mathbf{s} \mid \mathbf{s} = \gamma \mathbf{r}_m, \gamma \in \mathbb{R}^{*+}\}$  where  $\mathbf{r}_m \in \mathbb{R}^{L \times 1}$  is the  $m^{\text{th}}$  endmember,  $L$  the number of spectral bands. This model assumes that intra-class variability is only due to illumination variations for instance because of landscape slopes variation in non flat areas. However this model does not take into account other types of variability previously described: material can be weathered, composition variation, etc.... Variability can also be modelled as a statistical distribution when Bayesian approaches are used to perform the unmixing [8]. These models are highly parametric and require some strong assumptions. Other models describe variability as a bundle [17]. In the latter two cases the variability is not described by a factor,

the models lead to a more complex representation of the variability.

An experimental characterisation of intra-class variability is needed to develop a mixing model and the associated unmixing method. In [18], a brief characterisation of this variability is made. However this study is limited to two classes of well-defined materials and aims at showing the benefits of spectra standardisation regarding the shadow processing.

### B. Data description

Our spectral variability study is performed on an urban image. The studied area is located in the city of Toulouse, France. Toulouse city is composed of a large number of characteristic urban areas. Its downtown is typical from old cities in the South of France: tile roofs, low rise homes or big architectural monuments (cathedral, town hall...), very dense urbanisation, small streets or large avenues, green spaces (public gardens, squares...),... The suburbs are either residential (one-storey private houses or residential buildings) or industrial (large industrial buildings). In this study, we focus on images of the city center.

The airborne campaign was carried out in October 2013 [19]. The hyperspectral instrument is composed of two cameras: the first one covering the Visible Near Infra-Red (VNIR) domain (414nm to 992nm) with a 0.8m ground sampling distance (GSD) and the second one in the Short Wave Infra-Red (SWIR) range (980nm to 2498nm), with 1.8m GSD. To work on similar GSD, the VNIR image is degraded to 1.8m GSD and then registered with the SWIR image. It provides 405 bands over the range [414, 2498] nm with a 1.8m spatial resolution. An atmospheric compensation is applied to the data using the COCHISE tool [20]. After applying COCHISE, spectra are reflectance ones. After removing the atmospheric absorption bands (water vapor, CO<sub>2</sub>,...), the studied image contains 214 spectral bands. Figure 1 depicts some of these spectra.

Spectra were extracted from a portion of the entire image. It was chosen by considering the presence of large homogeneous areas, i.e. areas entirely covered by a supposedly “pure” material. Figure 2 shows the selected sub-image. The roof of the cathedral, at the center of the image, is particularly interesting since it is covered by various tiles with various slopes. In this sub-image, various areas with similar materials were selected, considered as a class and then the corresponding spectra were manually collected by only keeping those which meet the following criteria:

- pure material spectra
- various illumination conditions
- representative of the spectral variability of the materials due to composition, weathering (tiles of various roofs, asphalt extracted in different places...).

Three main classes of materials are considered: tiles, vegetation and asphalt. For these three classes, a large number of pure spectra can be extracted. It allowed relevant statistical studies.



Fig. 2. Sub-image selected for the intra-class variability study.

### C. Data analysis

To characterise the intra-class variability, two measures were computed onto the extracted spectra: the correlation between spectra and the projection on the first Principal Components. Let  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T$  denote the matrix composed of a set of extracted spectra, with  $\mathbf{y}_n \in \mathbb{R}^{L \times 1}$  a spectrum. The coefficients of the correlation matrix,  $CORR_{i,j}$ , can be written as follows:

$$CORR_{i,j} = \frac{Cov(\mathbf{y}_i, \mathbf{y}_j)}{\sigma_{y_i} \sigma_{y_j}}$$

with  $Cov(\cdot, \cdot)$  the covariance between two vectors and  $\sigma_y$  the standard deviation of the vector  $\mathbf{y}$ .

Figure 3 depicts four correlation matrices. For each of the tree and asphalt correlation matrices (Fig. 3 (c) and (d)), a single area of the sub-image corresponding to a single class was selected. The correlation was computed between all extracted spectra of this area. The tile correlation matrix (Fig. 3 (a)) was obtained in the same way, with tiles extracted from a same roof. The extended tile correlation matrix (Fig. 3 (b)) was computed between spectra extracted in various areas of the image, including the spectra used in the tile correlation matrix.

The correlation between the spectra extracted in a same roof fluctuates a lot, between 75% and 95%, which shows the large variability of the class. The low correlation between some spectra (under 75%) is probably due to the non purity of some extracted spectra, resulting from various elements that can be fixed on roofs (gutter, aerial...). These extreme values were removed for the following studies. Analyses of the asphalt correlation matrix are similar to the tile correlation matrix, the correlation between neighbouring asphalt spectra can be lower than 80%. The variability between spectra is due to different road surfacing. Results of the tree correlation matrices differ from the above results. The correlation between these spectra is higher than 90% (which correspond to a low variability) except for the correlation of several spectra with the 8<sup>th</sup> spectrum (probably a non pure spectrum). It can be due to the fact that in this area of Toulouse, most of the trees in streets belong to the same species, plane trees. The extended tile correlation matrix, in Fig. 3 (b), was obtained by computing pure spectra of a same material class extracted from

various locations of the image. It appears that the correlation between spectra extracted from a same roof is higher, from 75% to 95%, than the one between spectra of various roofs, from 65% to 95%. The largest observed intra-class variability, in Fig. 3 and Fig. 4, makes sense as the tiles of Toulouse roofs are not the same (various mineral compositions...). It means that intra-class variability increases when the number of building, or independent areas, increases. This result can be extended to other classes of material like asphalt.

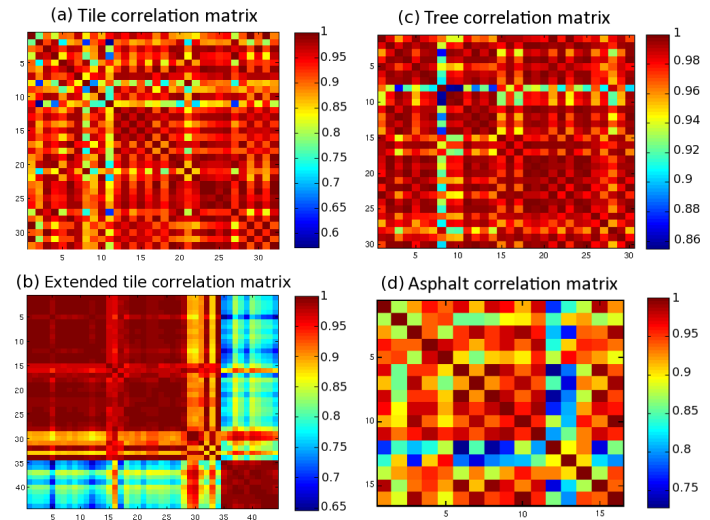


Fig. 3. Pixel-to-pixel correlation matrices of the 3 classes (tile, tree and asphalt) showing the intra-class variability.

Figure 4 depicts the correlation matrix of all extracted spectra which illustrates both inter and intra-class variabilities. Figure 4 shows the variability at the sub-image scale, since spectra from the same material were extracted from various locations of the sub-image, as for the extended tile correlation matrix in Fig. 3. Firstly it appears, as depicted in Figure 3, that the intra-class variability can be low (correlation higher than 80% for the asphalt class). However it is higher than in Fig. 3 for some classes, especially when the class includes spectra extracted in various areas of the images, like tiles. The intra-class variability of a class depends on the observed scene. The tree class frequently has a high intra-class variability. It is not the case here (correlation around 90%) since the considered areas contain similar tree species. The second important point is the high correlation between spectra of different classes. This variability is called inter-class variability. This phenomenon has to be considered. Indeed, if some classes have too close spectra, they would be hardly distinguishable especially when those classes have a high intra-class variability. For instance the correlation between some tile and vegetation spectra is higher than 60% whereas some tile spectra are only 70% correlated. This could lead to confusion between classes. Indeed if a class is defined by a bundle as in [17] the parameters have to be very carefully chosen to avoid spectra misclassification.

The above correlation coefficients do not depict the variability of average level from one spectrum to another, that is highly related to illumination variation. To visualize it, a

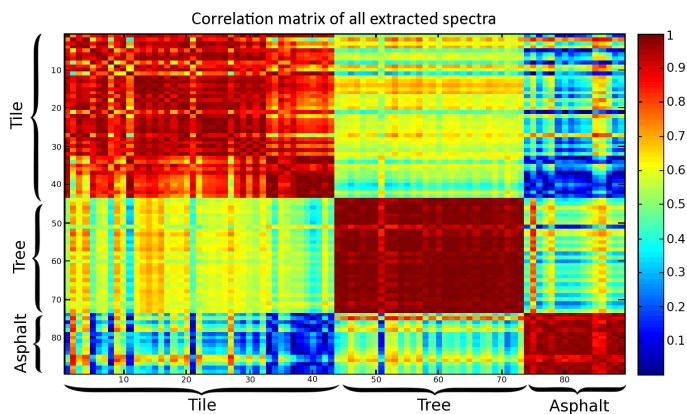


Fig. 4. Correlation matrices of all extracted spectra.

projection onto the first PCA axes was performed. Figure 5 illustrates the obtained results. In Figure 5 (a) spectra of three classes are projected onto the first two principal axes. For the tile class, spectra are extracted from various roof slopes (i.e. various illumination conditions). The tree class is composed of spectra extracted from a same area, however the illumination of the canopy varies from a tree to another. This class also depicts the effects of illumination on the intra-class variability. The scale factor variability which cannot be seen in Figures 3 and 4 appears clearly in Figure 5. Indeed the set of spectra  $\mathcal{E}_m = \{\mathbf{s} \mid \mathbf{s} = \gamma \mathbf{r}_m, \gamma \in \mathbb{R}^{*+}\}$  previously mentioned in subsection II-A is represented in a PCA projection as a line passing through the origin. In Figure 5 (b) this scale factor is visible for the sunny tile class. Spectra belonging to each class are more or less located along the same direction. However they do not exactly form a line. The variations on the axis passing through the origin correspond to the scale factor whereas the variations around this axis are linked to other phenomena. The location of blue and red dots in Figure 5 illustrates this. Indeed the darkest spectra (red dots) are the closest to the origin and the blue ones are situated further on the axis depending on their illumination. Asphalt projections corroborate this analysis. These spectra are dark ones and are located close to the origin and close to the spectra of dark trees and tiles.

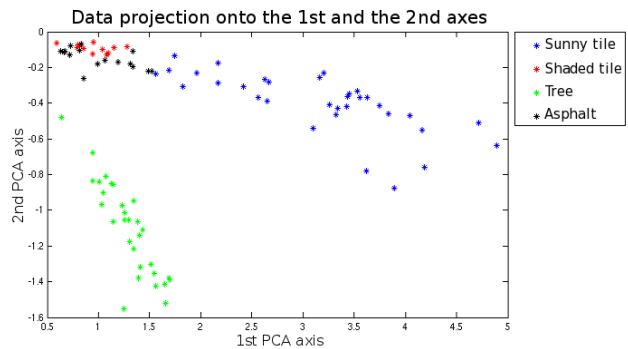
Taking these observations into account, it appears that the intra-class variability has to be considered to perform an accurate unmixing. These investigations also show that the modelling of the intra-class variability by a scale factor is not sufficient. So, in the following sections, a new mixing model dealing with intra-class variability is developed as well as unmixing methods adapted to this model.

### III. UNMIXING PROBLEM STATEMENT

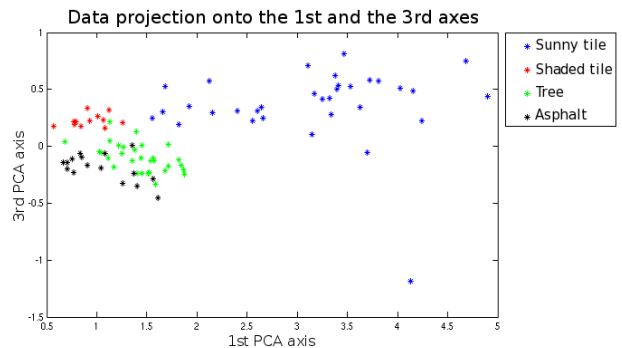
In the standard linear mixing model (LMM), each sensed spectrum  $\mathbf{x}_p \in \mathbb{R}^{L \times 1}$  can be written as follows:

$$\mathbf{x}_p = \sum_{m=1}^M c_{pm} \mathbf{r}_m \quad \forall p \in \{1, \dots, P\} \quad (1)$$

where  $p$  is the pixel index,  $P$  the number of pixels,  $m$  the index of one of the  $M$  endmembers,  $\mathbf{r}_m \in \mathbb{R}^{L \times 1}$  is the  $m^{\text{th}}$  source



(a) Projection onto the 1<sup>st</sup> and 2<sup>nd</sup> axes



(b) Projection onto the 1<sup>st</sup> and 3<sup>rd</sup> axes

Fig. 5. Projection on the first PCA axes of 3 extracted spectra classes: tile (blue and red dots), tree (green dots), asphalt (black dots).

spectrum and  $c_{pm}$  is the associated mixing coefficient. The above model does not take into account intra-class variability. We introduce a new extended mixing model, which can be rewritten as:

$$\mathbf{x}_p = \sum_{m=1}^M c_{pm} \mathbf{r}_m(p) \quad \forall p \in \{1, \dots, P\} \quad (2)$$

where  $\mathbf{r}_m(p)$  is the spectrum associated with the  $m^{\text{th}}$  source and the pixel  $p$ . Extracting these sources from such observations is an ill-posed problem. In Eq. (1) and (2) we add the classical sum-to-one constraint (besides the nonnegativity constraint). This condition leads to:

$$\sum_{m=1}^M c_{pm} = 1 \quad \forall p \in \{1, \dots, P\}. \quad (3)$$

Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_P]^T$  denote the sensed spectrum matrix,  $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_M]^T$  the endmember matrix if there is no intra-class variability (same sources for all pixels) and  $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_P]^T$  the mixing coefficient matrix. For all pixels  $p$ ,  $\mathbf{c}_p = [c_{p1}, \dots, c_{pM}]^T$  is an  $M$ -element vector containing the set of mixing coefficients associated with the  $p^{\text{th}}$  observed spectrum. The number of sources,  $M$ , is assumed to be known in the rest of the paper. The LMM can be written as follows:

$$\mathbf{X} = \mathbf{C}\mathbf{R}. \quad (4)$$

To obtain a similar expression from model (2), we introduce  $\mathbf{R}(p) = [\mathbf{r}_1(p), \dots, \mathbf{r}_M(p)]^T$ , the set of  $M$  sources associated

with the observed spectrum  $\mathbf{x}_p$ ,  $\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R}(1) \\ \dots \\ \mathbf{R}(P) \end{bmatrix} \in \mathbb{R}^{PM \times L}$ , the matrix containing all the sources and  $\tilde{\mathbf{C}} \in \mathbb{R}^{P \times PM}$  a block-diagonal matrix denoting the new mixing coefficient matrix:

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_1^T & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & \mathbf{c}_2^T & \dots & 0 \dots 0 \\ & & \ddots & \\ 0 \dots 0 & 0 \dots 0 & \dots & \mathbf{c}_P^T \end{bmatrix}. \quad (5)$$

So, Eq. (2) yields the matrix expression:

$$\mathbf{X} = \tilde{\mathbf{C}}\tilde{\mathbf{R}}. \quad (6)$$

We aim at retrieving matrices  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{R}}$  under the nonnegativity and sum-to-one constraints.

#### IV. UNCONSTRAINED PIXEL-BY-PIXEL NONNEGATIVE MATRIX FACTORISATION (UP-NMF)

Nonnegative Matrix Factorisation (NMF) has been adapted to solve remote sensing unmixing problems corresponding to the LMM [21], [22]. NMF aims at decomposing a matrix in a product of two nonnegative matrices, the coefficient matrix,  $\mathbf{C}$  and the reflectance matrix,  $\mathbf{R}$ . The necessary assumption of this method is the nonnegativity of the two searched matrices. The sum-to-one constraint can be added to NMF. The method can, then, be applied to the observation matrix. To perform NMF a cost function, hereafter called  $J_{nmf}$ , is minimised, however the obtained minima may be local ones. This point can be overcome with a good initialisation. This will be discussed in a later part (Sec. VI).

Now considering the extended mixing model (6), we introduce two evolutions of NMF for estimating  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{R}}$ . These two methods are respectively defined in this section and in Section V.

##### A. Cost function

Standard NMF aims at minimising the reconstruction error (RE) to extract a set of endmembers from a set of observed spectra. The standard cost function can be written as:

$$J_{nmf} = \frac{1}{2} \|\mathbf{X} - \mathbf{C}\mathbf{R}\|_F^2 \quad (7)$$

where  $\|\cdot\|_F$  is the Frobenius norm. In (7),  $\mathbf{C}$  and  $\mathbf{R}$  represent the *adaptive* variables which respectively aim at estimating the *actual* coefficients and source spectra involved in the mixing model (4). By using  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{R}}$  instead of  $\mathbf{C}$  and  $\mathbf{R}$  we extract one set of sources from each sensed spectrum. The Unconstrained pixel-by-pixel NMF (UP-NMF) method is an evolution of standard NMF which minimises this new cost function:

$$J_{upnmf} = \frac{1}{2} \|\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}}\|_F^2. \quad (8)$$

As in (7),  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{R}}$  are *adaptive* variables which aim at estimating the *actual* coefficients and source spectra in the mixing model (6).

##### B. Gradient calculation

To minimise  $J_{upnmf}$ , we developed an extended version of Lin's standard NMF algorithm [23]. Lin's method is based on projected gradient. So, we have to determine the derivatives of  $J_{upnmf}$  with respect to  $\tilde{\mathbf{R}}$  and  $\tilde{\mathbf{C}}$ . These derivatives are easily calculated with the matrix formulas in [24].

$$J_{upnmf} = \frac{1}{2} \text{Tr}((\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}})(\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}})^T) \\ \Rightarrow \frac{\partial J_{upnmf}}{\partial \tilde{\mathbf{R}}} = -\tilde{\mathbf{C}}^T(\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}}) \quad (9)$$

$$\frac{\partial J_{upnmf}}{\partial \tilde{\mathbf{C}}} = -(\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}})\tilde{\mathbf{R}}^T. \quad (10)$$

##### C. Update algorithm

The above calculations allow us to formulate the UP-NMF update algorithm. We choose to use a gradient descent algorithm. The update rules for  $\tilde{\mathbf{R}}$  and  $\tilde{\mathbf{C}}$  are the following:

$$\tilde{\mathbf{R}}^{(i+1)} \leftarrow \tilde{\mathbf{R}}^{(i)} - \alpha_{\tilde{\mathbf{R}}} \frac{\partial J_{upnmf}^{(i)}}{\partial \tilde{\mathbf{R}}} \quad (11)$$

$$\tilde{\mathbf{C}}^{(i+1)} \leftarrow \tilde{\mathbf{C}}^{(i)} - \alpha_{\tilde{\mathbf{C}}} \frac{\partial J_{upnmf}^{(i)}}{\partial \tilde{\mathbf{C}}} \quad (12)$$

followed by projection onto  $\mathbb{R}^{+*}$  and sum-to-one normalisation.  $\alpha_{\tilde{\mathbf{R}}}$  and  $\alpha_{\tilde{\mathbf{C}}}$  are positive adaptation steps.

Using (9) and (10), we obtain the update rules for our UP-NMF method. However only all  $\mathbf{c}_p^T$  in (5) should thus be updated, whereas the other terms of  $\tilde{\mathbf{C}}$  are kept to zero. So, the complete UP-NMF update can be written as follows:

1. Update of matrix  $\tilde{\mathbf{R}}$   
 $\tilde{\mathbf{R}}^{(i+1)} \leftarrow \tilde{\mathbf{R}}^{(i)} + \alpha_{\tilde{\mathbf{R}}}(\tilde{\mathbf{C}}^{(i)T}(\mathbf{X} - \tilde{\mathbf{C}}^{(i)}\tilde{\mathbf{R}}^{(i)}))$   
 $\tilde{\mathbf{R}}^{(i+1)} = \max(\tilde{\mathbf{R}}^{(i+1)}, \epsilon)$
2. Update of matrix  $\tilde{\mathbf{C}}$   
**for**  $p = 1$  **to**  $P$  **do**  
 $\mathbf{c}_p^{(i+1)T} \leftarrow \mathbf{c}_p^{(i)T}$   
 $\quad + \alpha_{\tilde{\mathbf{C}}}(\mathbf{x}_p^T - \mathbf{c}_p^{(i)T}\mathbf{R}(p)^{(i)})\mathbf{R}(p)^{(i)T}$   
 $\mathbf{c}_p^{(i+1)} = \max(\mathbf{c}_p^{(i+1)}, \epsilon)$   
**end**
3. Normalisation of the coefficients  
**for**  $p = 1$  **to**  $P$  **do**  
 $\mathbf{c}_p^{(i+1)} \leftarrow \mathbf{c}_p^{(i+1)} / \sum_{m=1}^M c_{pm}$   
**end**  
 where  $\epsilon$  is a small positive constant.

Due to the high underdeterminacy of this optimisation problem, the behavior of UP-NMF is not accurate enough. Estimated spectra  $\mathbf{r}_m(p)$  from a same class  $m$  may evolve so differently that they tend to define several classes of materials. This observation is all the more important that the inter-class variability is small, as it was explained in Sec. II-C. To limit this spreading of estimated spectra from a same class, constraints are required in the cost function. Such a constraint is proposed hereafter, Sec. V. It is based on the observations made in Sec. II-C and aims at penalising the spreading of estimated sources from the same class.

## V. INERTIA-CONSTRAINED PIXEL-BY-PIXEL NONNEGATIVE MATRIX FACTORISATION (IP-NMF)

### A. Cost function

Our second method is based on limiting class inertia to reduce the risk for an estimated pure spectrum to go out of its own class. This limitation is introduced in the optimisation problem by adding a penalty term in the cost function. The function  $J_{upnmf}$  to be minimised thus becomes:

$$J_{ipnmf} = \frac{1}{2} \|\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}}\|_F^2 + \mu \sum_{m=1}^M Tr(Cov(\tilde{\mathbf{R}}_{C_m})) \quad (13)$$

where  $\tilde{\mathbf{R}}_{C_m} \in \mathbb{R}^{P \times L}$  is the matrix containing estimates of all the endmembers of the  $m^{th}$  class (i.e. spectra of the  $m^{th}$  material),  $Tr(Cov(\tilde{\mathbf{R}}_{C_m}))$  is the inertia of the  $m^{th}$  class and  $\mu$  the constraint parameter. We compute  $\sum_{m=1}^M Tr(Cov(\tilde{\mathbf{R}}_{C_m}))$  rather than  $Tr(Cov(\tilde{\mathbf{R}}))$  because the latter expression tends to aggregate all classes. The covariance matrix trace measures the spreading of the spectra constituting the matrix, whatever the gravity center is. By using this penalty term, spectra moving too far away from others penalise all the class. Thus homogeneity is preserved. So it can be expected that the resulting spectra which evolved from the same ‘‘seed’’ still define the same class.

### B. Gradient calculation

To minimise  $J_{ipnmf}$ , we develop an extended version of Lin’s standard NMF algorithm [23], as for UPNMF. So, here again, we have to determine the derivatives of the cost function with respect to  $\tilde{\mathbf{R}}$  and  $\tilde{\mathbf{C}}$ . We start by decomposing  $J_{ipnmf}$  in two terms:

$$J_{ipnmf} = J_{RE} + \mu J_I \quad (14)$$

with

$$J_{RE} = \frac{1}{2} \|\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}}\|_F^2 \quad (15)$$

$$\begin{aligned} J_I &= \sum_{m=1}^M Tr(Cov(\tilde{\mathbf{R}}_{C_m})) \\ &= \sum_{m=1}^M \left( \frac{1}{P} Tr(\tilde{\mathbf{R}}_{C_m}^T \tilde{\mathbf{R}}_{C_m}) - \frac{1}{P^2} Tr(Q_{C_m}) \right) \end{aligned} \quad (16)$$

$$Q_{C_m} = \tilde{\mathbf{R}}_{C_m}^T \mathbf{1}_{P,P} \tilde{\mathbf{R}}_{C_m} \quad (17)$$

where  $\mathbf{1}_{P,P}$  is a  $P \times P$  dimensional matrix of ones.  $J_{RE}$  defines the reconstruction error (RE) and  $J_I$  the Inertia constraint. It can be noted that  $J_{RE}$  is equal to  $J_{upnmf}$ , so the calculations for this term have already been made.  $J_I$  does not depend on  $\tilde{\mathbf{C}}$  so the corresponding derivative is zero. To obtain the derivative of  $J_I$  with respect to  $\tilde{\mathbf{R}}$  we have to make

use of scalar writing. To this end, we start by rewriting  $J_I$  in order to extract  $\tilde{\mathbf{R}}$ . Eq. (16) yields:

$$\begin{aligned} J_I &= \sum_{m=1}^M \left( \frac{1}{P} \sum_{l=1}^L \sum_{k=1}^P [\tilde{R}_{C_m}]_{k,l}^2 - \frac{1}{P^2} Tr(Q_{C_m}) \right) \\ &= \frac{1}{P} \sum_{m=1}^M \sum_{l=1}^L \sum_{k=1}^P [\tilde{R}]_{(k-1)M+m,l}^2 - \frac{1}{P^2} \sum_{m=1}^M Tr(Q_{C_m}) \\ &= \frac{1}{P} \sum_{\kappa=1}^{PM} \sum_{l=1}^L [\tilde{R}]_{\kappa,l}^2 - \frac{1}{P^2} \sum_{m=1}^M Tr(Q_{C_m}) \\ &= \frac{1}{P} Tr(\tilde{\mathbf{R}}^T \tilde{\mathbf{R}}) - \frac{1}{P^2} \sum_{m=1}^M Tr(Q_{C_m}). \end{aligned} \quad (18)$$

We note that the first term can be derived by using the matrix formulas in [24]:

$$\frac{\partial}{\partial \tilde{\mathbf{R}}} \left( \frac{1}{P} Tr(\tilde{\mathbf{R}}^T \tilde{\mathbf{R}}) \right) = \frac{2}{P} \tilde{\mathbf{R}}. \quad (19)$$

Now focusing on the second term of Eq. (18), we introduce:

$$\begin{aligned} \mathbf{A} &= \tilde{\mathbf{R}}_{C_m}^T \longrightarrow a_{ij} = [\tilde{R}_{C_m}]_{ji} \\ \mathbf{B} &= \mathbf{1}_{P,P} \tilde{\mathbf{R}}_{C_m} \longrightarrow b_{ij} = \sum_{\beta=1}^P [\tilde{R}_{C_m}]_{\beta j}. \end{aligned}$$

Then  $Q_{C_m} = \mathbf{A}\mathbf{B}$  and

$$\begin{aligned} [Q_{C_m}]_{ij} &= \sum_{\alpha=1}^P a_{i\alpha} b_{\alpha j} \\ &= \sum_{\alpha=1}^P \sum_{\beta=1}^P [\tilde{R}_{C_m}]_{\alpha i} [\tilde{R}_{C_m}]_{\beta j}. \end{aligned}$$

Therefore:

$$\begin{aligned} Tr(Q_{C_m}) &= \sum_{l=1}^L [Q_{C_m}]_{l,l} \\ &= \sum_{l=1}^L \sum_{\alpha=1}^P \sum_{\beta=1}^P [\tilde{R}_{C_m}]_{\alpha l} [\tilde{R}_{C_m}]_{\beta l}. \end{aligned} \quad (20)$$

From Eq. (20) we can calculate the derivative of  $\sum_{m=1}^M Tr(Q_{C_m})$  with respect to  $[\tilde{R}]_{\gamma\lambda}$ , one element of  $\tilde{\mathbf{R}}$ . By definition of  $\tilde{\mathbf{R}}_{C_m}$ ,  $[\tilde{R}]_{\gamma\lambda}$  is present in only one of the matrices  $Q_{C_m}$ , i.e. the one with  $m = 1 + (\gamma - 1)(\text{mod } M)$ , denoted as  $\eta$  hereafter.

$$\begin{aligned} \frac{\partial}{\partial [\tilde{R}]_{\gamma\lambda}} \left( \sum_{m=1}^M Tr(Q_{C_m}) \right) &= \sum_{m=1}^M \frac{\partial}{\partial [\tilde{R}]_{\gamma\lambda}} (Tr(Q_{C_m})) \\ &= \frac{\partial}{\partial [\tilde{R}]_{\gamma\lambda}} (Tr(Q_{C_\eta})) \\ &= \frac{\partial}{\partial [\tilde{R}]_{\gamma\lambda}} \left( \sum_{l=1}^L \sum_{\alpha=1}^P \sum_{\beta=1}^P [\tilde{R}_{C_\eta}]_{\alpha l} [\tilde{R}_{C_\eta}]_{\beta l} \right). \end{aligned} \quad (21)$$

From Eq. (21), four cases are possible:

$$\frac{\partial(\sum_{l=1}^L [\tilde{R}_{c_\eta}]_{\alpha l} [\tilde{R}_{c_\eta}]_{\beta l})}{\partial [\tilde{R}]_{\gamma \lambda}} = \begin{cases} 0 & \text{if } \alpha \neq \gamma \text{ and } \beta \neq \gamma, \\ [\tilde{R}_{c_\eta}]_{\beta \lambda} & \text{if } \alpha = \gamma \text{ and } \beta \neq \gamma, \\ [\tilde{R}_{c_\eta}]_{\alpha \lambda} & \text{if } \alpha \neq \gamma \text{ and } \beta = \gamma, \\ 2[\tilde{R}_{c_\eta}]_{\alpha \lambda} & \text{if } \alpha = \beta = \gamma, \end{cases}$$

therefore:

$$\begin{aligned} \frac{\partial}{\partial [\tilde{R}]_{\gamma \lambda}} \left( \sum_{m=1}^M \text{Tr}(Q_{c_m}) \right) &= 2 \sum_{\alpha=1}^P [\tilde{R}_{c_\eta}]_{\alpha \lambda} \\ &= 2 \sum_{\alpha=1}^P [\tilde{R}]_{(\alpha-1)M+\eta, \lambda}. \end{aligned} \quad (22)$$

Result (22) can be extended to all entries of  $\tilde{\mathbf{R}}$ , which yields:

$$\frac{\partial}{\partial \tilde{\mathbf{R}}} \left( \sum_{m=1}^M \text{Tr}(Q_{c_m}) \right) = 2\mathbf{U}\tilde{\mathbf{R}} \quad (23)$$

with  $\mathbf{U} \in \mathbb{R}^{PM \times PM}$  defined by

$$\mathbf{U} = \begin{bmatrix} & \overbrace{\begin{matrix} 1 & 0 & \dots & 0 & 1 & \dots \\ 0 & 1 & \dots & 0 & 0 & \dots \\ \vdots & & \ddots & & \vdots & \\ 0 & 0 & \dots & 1 & 0 & \dots \\ 1 & 0 & \dots & 0 & 1 & \dots \\ \vdots & & & & & \ddots \end{matrix}}^M \\ M & \left[ \begin{matrix} \mathbf{Id}_M & \dots & \mathbf{Id}_M \\ \vdots & \ddots & \vdots \\ \mathbf{Id}_M & \dots & \mathbf{Id}_M \end{matrix} \right] \end{bmatrix} = \begin{bmatrix} \mathbf{Id}_M & \dots & \mathbf{Id}_M \\ \vdots & \ddots & \vdots \\ \mathbf{Id}_M & \dots & \mathbf{Id}_M \end{bmatrix}.$$

The notation  $\mathbf{Id}_D$  stands for the  $D$ -dimensional identity matrix. Thanks to (18), (19) and (23) we obtain the partial derivatives of  $J_I$ :

$$\frac{\partial J_I}{\partial \tilde{\mathbf{R}}} = \frac{2}{P} (\mathbf{Id}_{PM} - \frac{1}{P} \mathbf{U}) \tilde{\mathbf{R}}. \quad (24)$$

By combining (9), (10) and (24) we obtain the two partial derivatives of the general cost function (14) with respect to  $\tilde{\mathbf{R}}$  and  $\tilde{\mathbf{C}}$ :

$$\begin{aligned} \frac{\partial J_{ipnmf}}{\partial \tilde{\mathbf{R}}} &= -\tilde{\mathbf{C}}^T (\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}}) + \frac{2\mu}{P} (\mathbf{Id}_{PM} - \frac{1}{P} \mathbf{U}) \tilde{\mathbf{R}} \\ \frac{\partial J_{ipnmf}}{\partial \tilde{\mathbf{C}}} &= -(\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{R}})\tilde{\mathbf{R}}^T. \end{aligned}$$

### C. Update algorithm

The previous calculations allow us to formulate the update algorithm. The general gradient descent formulation was already given in (11) and (12). Using it we obtain the update rules for our IP-NMF method. As for UP-NMF, only the  $\mathbf{c}_p^T$  in (5) should thus be updated, whereas the other terms of  $\tilde{\mathbf{C}}$  are kept to zero. So, the complete IP-NMF update can be written as follows:

$$\begin{aligned} &1. \text{ Update of matrix } \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}}^{(i+1)} &\leftarrow \tilde{\mathbf{R}}^{(i)} + \alpha_{\tilde{\mathbf{R}}} (\tilde{\mathbf{C}}^{(i)T} (\mathbf{X} - \tilde{\mathbf{C}}^{(i)} \tilde{\mathbf{R}}^{(i)}) \\ &\quad - \frac{2\mu}{P} (\mathbf{Id}_{PM} - \frac{1}{P} \mathbf{U}) \tilde{\mathbf{R}}^{(i)}) \\ \tilde{\mathbf{R}}^{(i+1)} &= \max(\tilde{\mathbf{R}}^{(i+1)}, \epsilon) \end{aligned}$$

### 2. Update of matrix $\tilde{\mathbf{C}}$

$$\begin{aligned} &\text{for } p = 1 \text{ to } P \text{ do} \\ &\quad \mathbf{c}_p^{(i+1)T} \leftarrow \mathbf{c}_p^{(i)T} \\ &\quad \quad + \alpha_{\tilde{\mathbf{C}}} (\mathbf{x}_p^T - \mathbf{c}_p^{(i)T} \mathbf{R}(p)^{(i)}) \mathbf{R}(p)^{(i)T} \\ &\quad \mathbf{c}_p^{(i+1)} = \max(\mathbf{c}_p^{(i+1)}, \epsilon) \\ &\text{end} \end{aligned}$$

### 3. Normalisation of the coefficients

$$\begin{aligned} &\text{for } p = 1 \text{ to } P \text{ do} \\ &\quad \mathbf{c}_p^{(i+1)} \leftarrow \mathbf{c}_p^{(i+1)} / \sum_{m=1}^M c_{pm} \\ &\text{end} \end{aligned}$$

As all iterative algorithms, IP-NMF as UP-NMF, have to be initialised. The choices we made are described in the following Section VI-A.

## VI. TEST RESULTS ON SEMI-SYNTHETIC DATA SET

### A. Test description

Tests of the UP-NMF and IP-NMF methods are firstly performed with semi-synthetic data. The sources are real spectra extracted from the hyperspectral sub-image described in Sec. II-B, they are depicted in Figure. 1. They describe realistic source variations. Mixing coefficients are randomly chosen while respecting the sum-to-one constraint. The mixing model is defined by Eq. (2).

Tests are performed by varying the initialisations  $\tilde{\mathbf{R}}^{(0)}$  and  $\tilde{\mathbf{C}}^{(0)}$  and the constraint parameter  $\mu$ . Three  $\tilde{\mathbf{R}}^{(0)}$  are tested: (i)  $M$  mixed signals randomly selected from the observations, (ii) the  $M$  purest signals extracted with a classical remote sensing blind source separation method (N-FINDR [25]), (iii) for each class  $m$ , the average of all  $P$  source signals in this class. Mixing coefficient initialisation is obtained in two ways: (a) by giving the same constant value,  $\frac{1}{M}$ , to all coefficients, (b) by extracting coefficients associated with initialised spectra with a Fully Constrained Least Square (FCLS) regression [26], [27], [28]. The constraint parameter  $\mu$  varies from 0 to 100 to assess the impact of  $\mu$  on the algorithm performance.

Among the above listed initialisations, only the results for one couple are discussed in the first part of Section VI-C, namely  $\tilde{\mathbf{R}}^{(0)}$  and  $\tilde{\mathbf{C}}^{(0)}$  respectively initialised by (ii) and (a). Indeed (i) is a more uncertain initialisation than (ii) and (iii) is the best expected initialisation of  $\tilde{\mathbf{R}}^{(0)}$  but it requires knowledge about the data which is not available for real data sets. (a) was chosen to initialise  $\tilde{\mathbf{C}}^{(0)}$  to reduce the risk of being in a local minimum (this conclusion results from a study of UP-NMF and IP-NMF initialised with (b)).

### B. Evaluation criteria

Chosen criteria assess the benefits of our method. So, a major point to be evaluated is the correspondence between estimated pure material reflectance spectra and spectra really present in each pixel. To this end we computed, in each pixel  $p$ , the spectral angles between these two sets of spectra [13]. The resulting criterion is defined as:

$$SAM(p) = \frac{1}{M} \sum_{m=1}^M \left( \cos^{-1} \left( \frac{\langle \mathbf{r}_m(p), \hat{\mathbf{r}}_m(p) \rangle}{\|\mathbf{r}_m(p)\|_2 \cdot \|\hat{\mathbf{r}}_m(p)\|_2} \right) \right)$$

where  $\langle \cdot, \cdot \rangle$  stands for the scalar product,  $\mathbf{r}_m(p)$  the spectrum of the  $m^{\text{th}}$  source really present in the  $p^{\text{th}}$  pixel and  $\hat{\mathbf{r}}_m(p)$  the estimated one. The spectral angle error  $SAM(p)$  is then averaged over all pixels to obtain the mean spectral error,  $SAM$ .

Two other criteria were also computed. The first one is the reconstruction error,  $RE$ , to evaluate the benefit of our method on the global reconstruction of the image. Like for the spectral error, the reconstruction error is computed at pixel level or averaged over all the image:

$$RE(p) = \frac{1}{L} \cdot \|\mathbf{x}_p - \hat{\mathbf{c}}_p^T \hat{\mathbf{R}}(p)\|_F.$$

The second one is the mean square error computed on the coefficients,  $CE$ . As for  $RE$  and  $SAM$  it can be computed at pixel level or averaged on all image:

$$CE(p) = \frac{1}{M} \cdot \|\mathbf{c}_p - \hat{\mathbf{c}}_p\|_F.$$

Computing these errors is possible for semi-synthetic data for which the sources and mixing parameters are known.

### C. Results

Figure 6 illustrates the sources used in the mixing (blue, red and green stars) compared with UP-NMF results (black, cyan and yellow stars) and standard NMF results (blue, red and green circles). Figure 7 and Figure 8 illustrate the same comparisons for IP-NMF with  $\mu$  equal to 30 and 100 respectively. For these three figures,  $\hat{\mathbf{R}}^{(0)}$  is built with N-FINDR results and  $\hat{\mathbf{C}}^{(0)}$  with constant coefficients  $\frac{1}{M}$ .

For each class, the scatter plots of the actual and extracted spectra should be superimposed up to scale factors. Indeed scale factors can be contained in the estimated spectra and respectively the inverse in the associated estimated abundances:

$$\mathbf{x}_p = \sum_{m=1}^M \frac{1}{k_m(p)} c_{pm} \times k_m(p) \mathbf{r}_m(p) \quad \forall p \in \{1, \dots, P\}. \quad (25)$$

And hence dots can move onto the axes passing through the origin. Thus the dot position onto its class main axis depends on its estimated scale factor.

As mentioned in Section IV, the main risk is spectra spreading. The magnitude of the spreading of a class can be measured by the variance of the projection of this class plot onto the line which contains the center of gravity of that class plot and which is orthogonal to the line which contains that center of gravity and the origin. This spreading clearly appears in Figure 6, the above defined variances for the retrieved spectra are high as compared with the variances for the spectra used in the mixing. This reinforces our idea to limit the class inertia. Results in Table I confirm this observation. In this table, results of UP-NMF and IP-NMF (with  $\mu$  fixed to 30 and 100) were compared to standard NMF and a classical unmixing method (N-FINDR [29] + FCLS [26]). The average SAM of the UP-NMF method is the highest due to the poor reconstruction of some spectra. However we can argue that the Reconstruction Error ( $RE$ ) is the best for this method. But, as for NMF, the  $RE$  of UP-NMF have to be carefully interpreted.

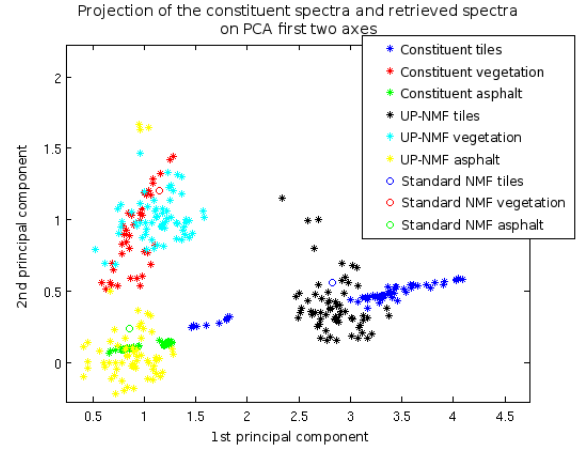


Fig. 6. Projection onto the first two PCA axes of constituent spectra (blue, red, green stars), UP-NMF spectra (black, cyan, yellow stars) and standard NMF spectra (blue, red, green circles).

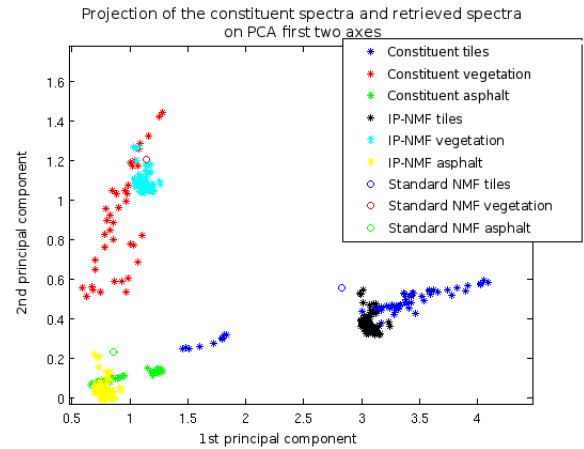


Fig. 7. Projection onto the first two PCA axes of constituent spectra (blue, red, green stars), IP-NMF spectra (black, cyan, yellow stars) with  $\mu = 30$  and standard NMF spectra (blue, red, green circles).

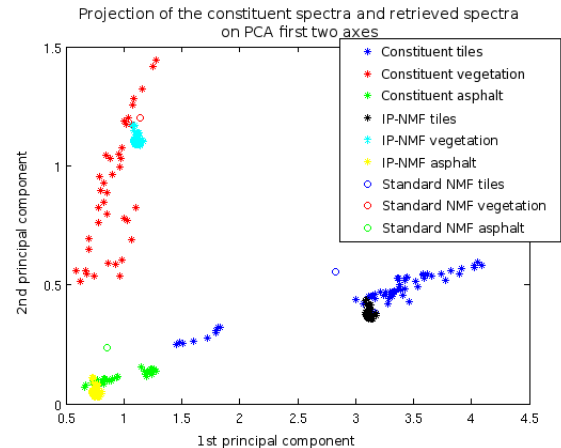


Fig. 8. Projection onto the first two PCA axes of constituent spectra (blue, red, green stars), IP-NMF spectra (black, cyan, yellow stars) with  $\mu = 100$  and standard NMF spectra (blue, red, green circles).



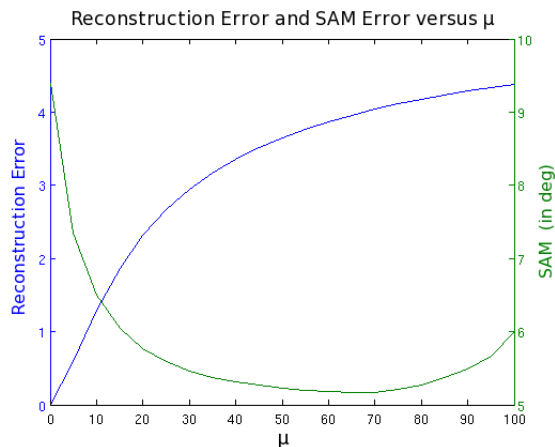


Fig. 9. Evolution of the RE (blue curve) and the SAM (green curve) as a function of  $\mu$  parameter

Indeed  $RE$  minimisation is what UP-NMF and standard NMF look for, as shown in (7) and (8). So we can obtain good  $RE$  and poor  $SAM$  and  $CE$  with these methods. In Figure 7 and Figure 8, the evolution of the scatter plot with respect to  $\mu$  shows the impact of the inertia constraint. If  $\mu$  is very high, the IP-NMF method gets closer to a standard unmixing method since it yields very compact sets of spectra for each class (close to a single spectrum). In the case when  $\mu$  is suitable, the widths (related to the above mentioned variance) of the scatter plots formed by the estimated spectra are comparable to those of the scatter plots formed by the constituent spectra. That is why  $\mu = 30$  is a suitable parameter value: it yields a relevant variance orthogonally to the main axis of the class. Since scale factors do not affect SAM, computing these SAM allows one to confirm this point. Table I shows improvements brought by IP-NMF in particular concerning the  $SAM$ . Indeed with IP-NMF an improvement of 2 degrees is observed. Results are given for only 3 values of  $\mu$  (0 or UP-NMF, 30 and 100). They were selected to describe three examples of visually different results. However, as shown by Figure 9, the evolutions of the  $RE$  and  $SAM$  are smooth with respect to  $\mu$ . That means that  $\mu$  is not a highly sensitive parameter. It would be possible to fix it to 50 and obtain similar results. Even  $\mu$  equal to 100 gives better results than the standard NMF and the classical method according to Table I. The computational times of UP-NMF and IP-NMF are significantly higher than those of these two standard methods. However the structure of the proposed algorithm makes it quite easily parallelisable. The time variations between the 3 cases are caused by the fluctuations of the number of iterations.

The second point consists in analysing the impact of initialisation. We applied IP-NMF with the initialisation scenarios described in Sec. VI-A. It appears that a poor initialisation (scenario (i)) leads to poor results for both the standard NMF, UP-NMF and IP-NMF. Yet, compared with standard NMF, IP-NMF improves the average spectral angle error in every initialisation case. We also noted that if both the spectra and the abundance coefficients are initialised too close to a local

TABLE I  
GLOBAL COMPARISON OF UNMIXING METHODS.

	N-FINDR +FCLS	Standard NMF	UP-NMF	IP-NMF ( $\mu = 30$ )	IP-NMF ( $\mu = 100$ )
$RE$	17.5	0.6	< 0.1	2.9	4.4
$SAM$ (deg)	7.7	7.7	9.4	5.5	6.1
$CE$ (%)	4.0	4.7	3.8	3.8	3.8
Time (sec)	0.03	0.02	5.87	5.90	6.48

minimum (scenario (iii) and (b)), the results of our methods are close to this initialisation.

## VII. TEST RESULTS FOR REAL IMAGE

### A. Data set

In this study we focus on city center image. A sub-image was extracted from the larger image depicted in Figure 2 and described in Sec. II-B. It contains an avenue, vegetation (trees and grass), tile roofs and shadows of these materials. Figure 10 depicts this sub-image.

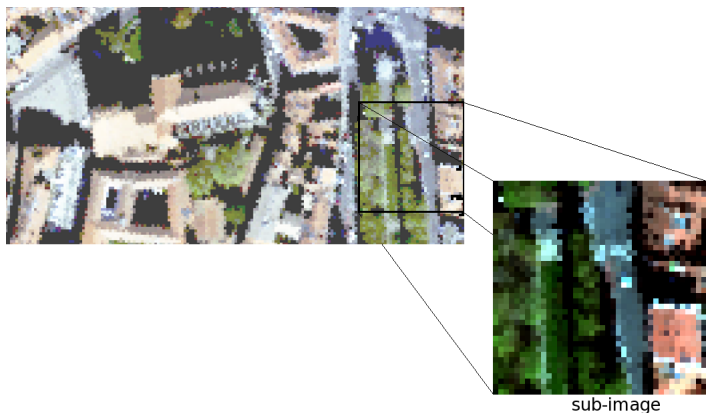


Fig. 10. Extracted sub-image.

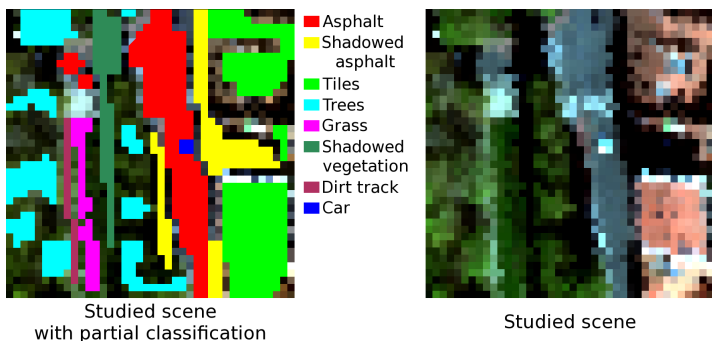


Fig. 11. A zoom of the studied scene (right hand side) in real colors ([647nm 542nm 454nm]) and a manual partial classification of this scene (left hand side). The manual partial classification allows a better understanding of the selected area.

## B. Test description

Several methods were tested on this data set:

- Two classical geometric methods: VCA [16] and N-FINDR [25] extract endmembers (one per class of materials). The abundances are then retrieved by a usual Fully Constrained Least Square (FCLS) algorithm [27], [28].
- A standard NMF method [23] is applied to retrieve at the same time both the source spectra (one per class) and their associated abundance coefficients.
- IP-NMF is applied to obtain one set of endmembers per pixel and the associated coefficients.

IP-NMF asks for the number of classes,  $M$ . To find it a well-known method, HySime, developed by Bioucas-Dias et al. in [30], was tested to identify the number of endmembers in the hyperspectral image. The results of HySime lead to 46 endmembers. This over-determination of the endmember extraction is due to the intra-class variability. Indeed HySime does not take this phenomenon into account (at least does not enough). The aim of IP-NMF is to work with a limited and realistic number of classes. So this kind of algorithm cannot be used to initialise  $M$ . It will be manually initialised. In these tests IP-NMF is applied with various parameter initialisations. Figure 11 shows that the observed scene mainly consists of three classes: asphalt (red and yellow classes in Fig. 11), vegetation (cyan, magenta and dark green in Fig. 11) and tiles (green class in Fig. 11). Some particular elements could be problematic (the car (blue class in Fig. 11), the track (purple class in Fig. 11), the roof elements (mixed in the tile pixels)...). So the number of endmembers,  $M$ , is successively fixed to 3, 5 and 7. In order to perform a fully automated algorithm the initialisation of spectra is firstly obtained by VCA. As a comparison, a second initialisation with manually selected spectra is also performed. Thanks to this manual selection of pure spectra we obtain one of the best possible initialisations. The coefficients are initialised with the constant value  $\frac{1}{M}$  since this initialisation was the one yielding the best results for the semi-synthetic data tests (cf. Subsection. VI-C). Considering Fig. 9, the  $SAM$  only moderately varies when  $\mu$  ranges between 20 and 80. So a constant  $\mu$  was fixed for all the following tests, equal to 30, as for the semi-synthetic data tests.

Evaluation criteria cannot be the same as those used for the semi-synthetic data test. Indeed due to the lack of ground truth we cannot compute the  $SAM$  and  $CE$  errors. Therefore only the Reconstruction error,  $RE$  and a usual analysis of the results in various ways (PCA projection, abundance maps...) can be considered. It was decided not to exploit the  $RE$  because of the reasons developed in Section VI-C.

## C. Results

IP-NMF was applied with the two initialisations described in the previous subsection. Results are depicted in Figures 12 to 14. Each of these figures shows the abundance maps of VCA + FCLS unmixing (1<sup>st</sup> line), NMF (2<sup>nd</sup> line), IP-NMF initialised with VCA (3<sup>rd</sup> line) and IP-NMF manually initialised (4<sup>th</sup> line) respectively applied with  $M$  equal to 3, 5 and 7. The results obtained with N-FINDR are not shown

because they are very close to the ones given by VCA.

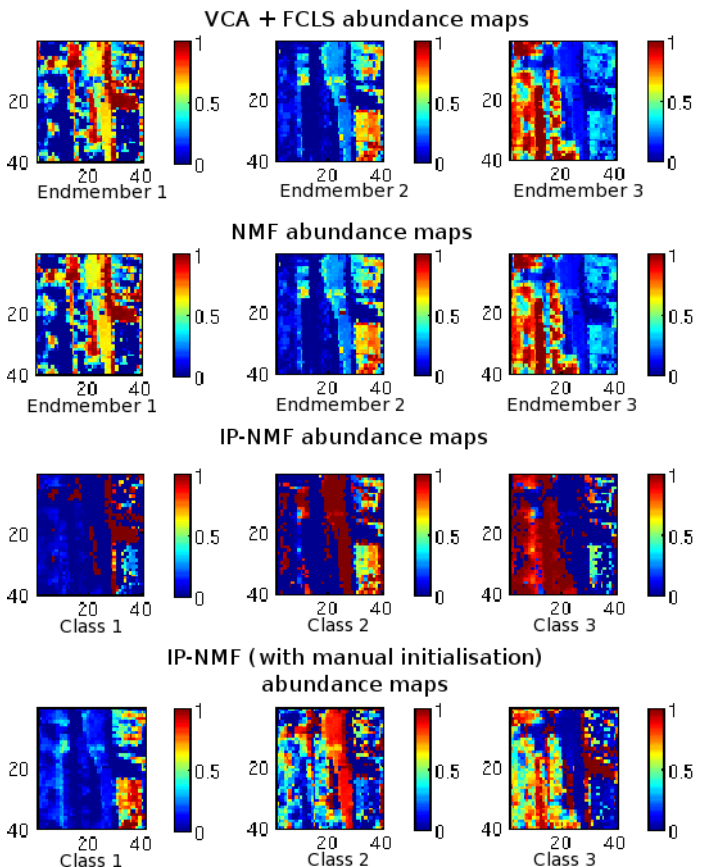


Fig. 12. Abundance maps obtained for  $M = 3$  with 4 methods: VCA + FCLS (1<sup>st</sup> line), standard NMF (2<sup>nd</sup> line), IP-NMF initialised with VCA (3<sup>rd</sup> line) and IP-NMF initialised with manually selected spectra (4<sup>th</sup> line).

1) *Results of classical unmixing methods (one set of endmembers per image):* As described previously, two methods were applied to the image, VCA + FCLS and a standard NMF. Abundance maps are depicted in the 1<sup>st</sup> and 2<sup>nd</sup> lines of Figures 12 to 14. It appears that VCA + FCLS and NMF maps are relatively similar, this is due to the initialisation. Indeed NMF was initialised by VCA resulting spectra. Several observations are made from these maps.

Firstly the asphalt class (road) is never accurately extracted. In the three figures it is a mixture between shadow endmember and another one. For instance in Figure 12 it is a mixture between tile endmember and shadow endmember. Then, increasing the number of endmembers reduces the unmixing quality because these unmixing methods then sometimes focus on specific spectra (vertices of the simplex), instead of extracting all major classes. It is the case in Figure 13 and Figure 14 where a car spectrum is extracted (2<sup>nd</sup> column and 1<sup>st</sup> column respectively in Fig. 13 and Fig. 14). Conversely due to intra-class variability several endmembers can be extracted belonging to the same class, for instance the tiles in Figure 14 (5<sup>th</sup> and 6<sup>th</sup> columns).

For this image it seems that Figure 12 depicts the best results. Indeed except the road all classes are depicted. However

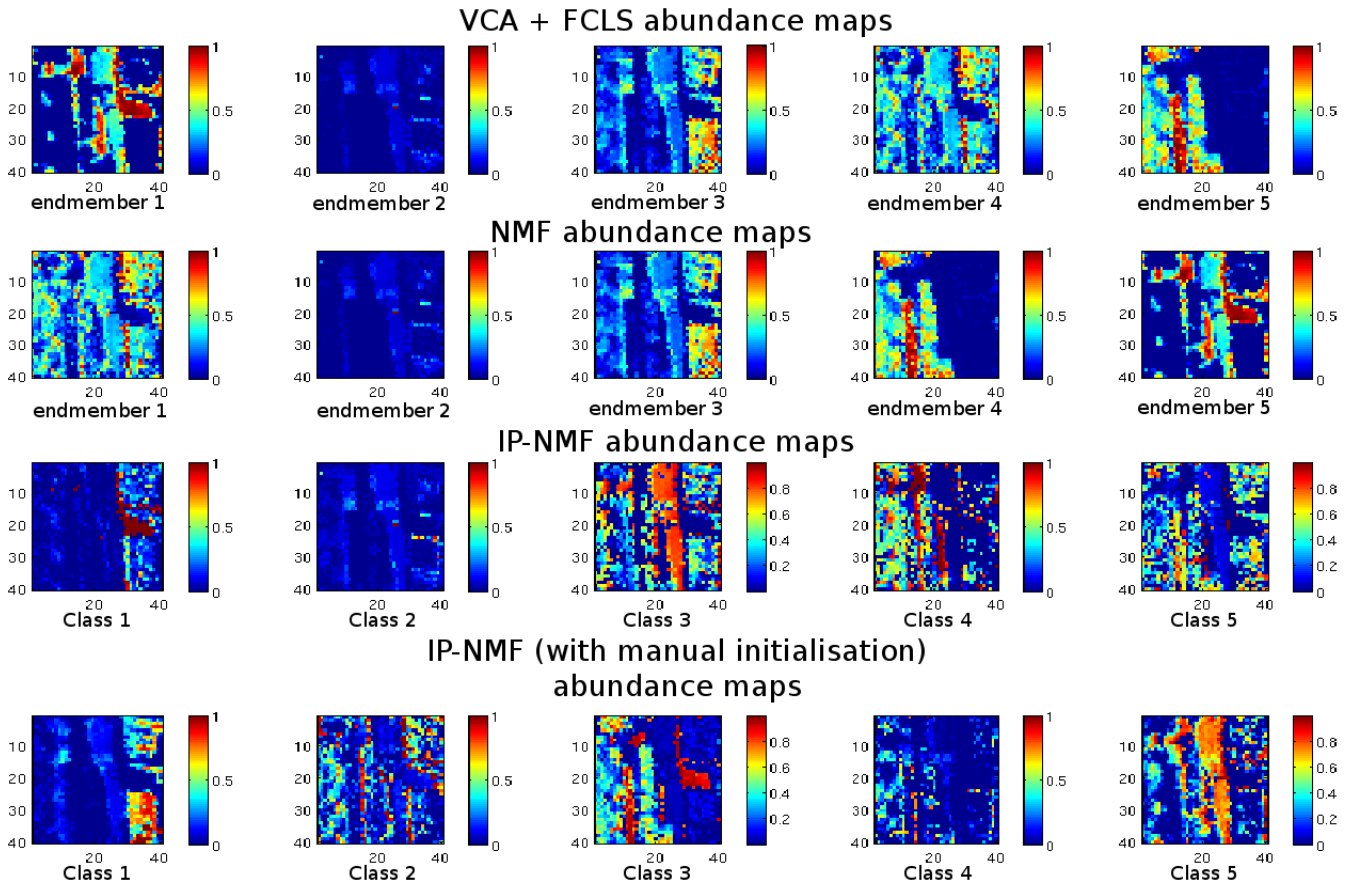


Fig. 13. Abundance maps obtained for  $M = 5$  with 4 methods: VCA + FCLS (1<sup>st</sup> line), standard NMF (2<sup>nd</sup> line), IP-NMF initialised with VCA (3<sup>rd</sup> line) and IP-NMF initialised with manually selected spectra (4<sup>th</sup> line).

the vegetation (3<sup>rd</sup> column) is not uniform. Indeed it is the *shadow* spectrum which was used to rebuild the shadowed tree spectra.

2) *Results of IP-NMF method initialised with VCA*: To analyse the “automated” IP-NMF results the 3<sup>rd</sup> line of Figures 12, 13 and 14 is studied. First of all it has to be noted that even if abundance maps are presented in the same way for VCA, NMF and IP-NMF, they depict, in fact, two different things. Indeed for VCA and NMF each map is associated with a single endmember whereas for IP-NMF a map is associated with a class. This means that each abundance of IP-NMF abundance maps is associated with a spectrum different from its neighbours. The understanding of the results is highly linked to this preliminary remark.

Our first observation concerning Figure 12 is the homogeneity of areas: contrary to the VCA and NMF results, there is no cyan background in the maps, which would correspond to a material weakly present in a large number of pixels. This is the result of the introduction of degrees of freedom for modelling intra-class variability in our adaptive variables corresponding to  $\tilde{C}$  and  $\tilde{R}$ . The three classes obtained with IP-NMF are: shadowed asphalt (1<sup>st</sup> column), sunny asphalt + tiles (2<sup>nd</sup> column), vegetation (3<sup>rd</sup> column). The 2<sup>nd</sup> class contains spectra of asphalt and tiles. This result is the consequence of two things: the low inter-class variability (cf. Sec. II-C) and

the initialisation. Indeed, due to the low inter-class variability, constituent spectra are spectrally close even if they do not belong to the same class. So it is possible that, in a same class, estimated spectra evolve in different ways and form two clusters, i.e. two sub-classes inside one class. This is what happens in this test. This is possible since the distance between the two center of gravity of the sub-classes is similar to the intra-class variability of some classes. Initialisation is also accountable for this phenomenon since all the spectra of a class evolved from a same “seed spectrum”. So, if the initialisation spectrum (or “seed spectrum”) is close to the two classes this phenomenon more easily appears.

Since this phenomenon is due to the combination of two classes which are spectrally too close, what happens when the number of classes is increased ? This is depicted in Figures 13 and 14. It can be observed that the results are not as expected. Indeed new classes appear, for instance the car (Fig. 13, 2<sup>nd</sup> column and Fig. 14, 1<sup>st</sup> column) but the quality of the unmixing decreases. For instance the road is now considered as a mixture of two spectra. This decrease of the quality is due to the initialisation, as will be shown in Section VII-C3. Indeed the VCA spectra are used to initialise IP-NMF and the poor quality of spectra extracted by VCA influences the results of IP-NMF, in which spectra cannot move enough to make up for the initialisation error. However

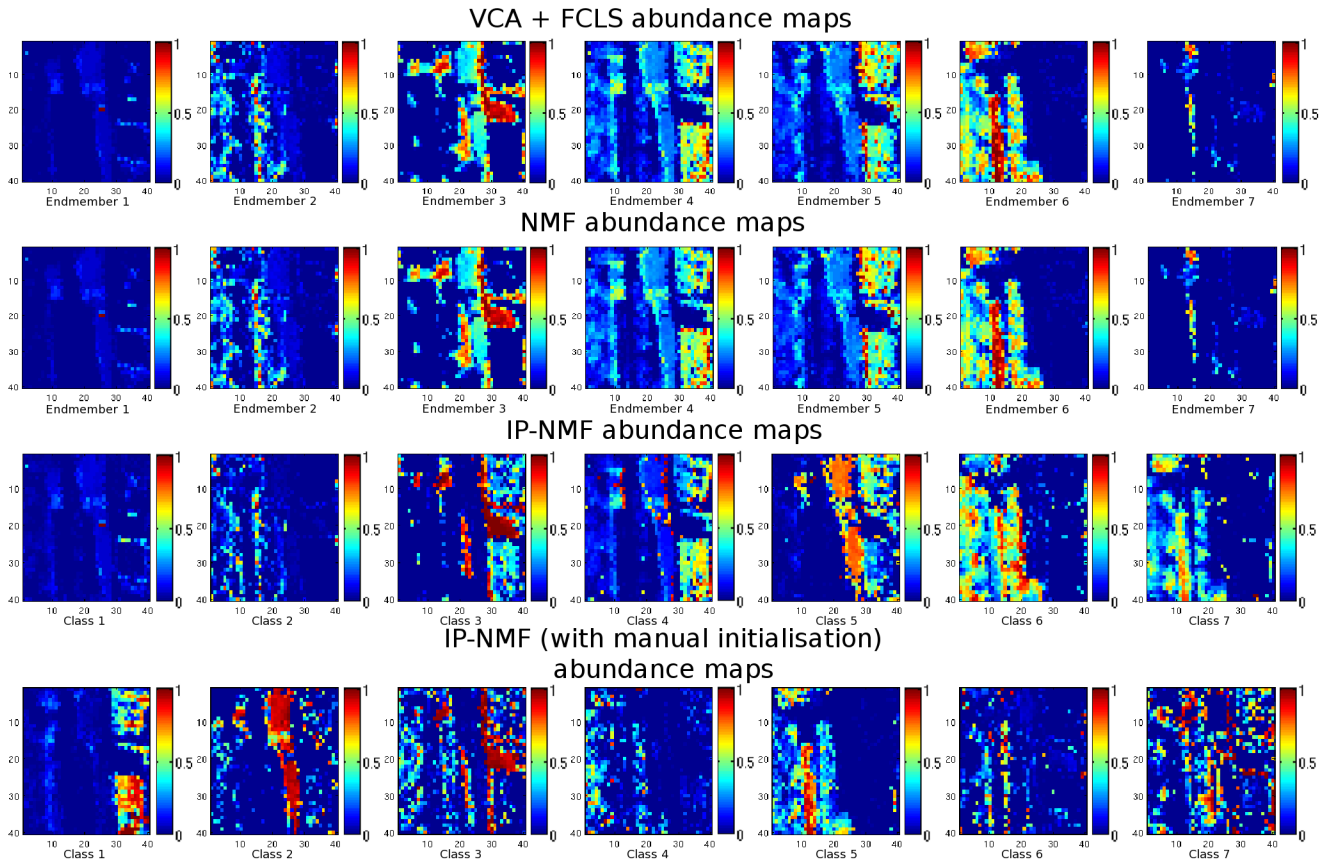


Fig. 14. Abundance maps obtained for  $M = 7$  with 4 methods: VCA + FCLS ( $1^{st}$  line), standard NMF ( $2^{nd}$  line), IP-NMF initialised with VCA ( $3^{rd}$  line) and IP-NMF initialised with manually selected spectra ( $4^{th}$  line).

IP-NMF gives better results than VCA or NMF. Indeed some classes are detected only by IP-NMF. Besides sometimes spectra in a pixel are misclassified by IP-NMF (i.e. they do not belong to the same physical class as other spectra of this class) but anyway they accurately fit the actual pure spectra present in this pixel.

### 3) Results of IP-NMF method with manual initialisation:

The  $4^{th}$  line of Figures 12, 13 and 14 shows the impact of the initialisation on IP-NMF results. IP-NMF was initialised respectively by 3, 5 and 7 spectra manually selected in the image. This allows one to choose the initial classes and ensures that initialisation spectra are pure ones. In Figure 12, selected spectra were a tile spectrum, an asphalt spectrum and a tree spectrum. Abundance maps show that the tiles are very well extracted (column 1 of Fig. 12 compared to the sub-image of Fig. 10). Asphalt is quite well extracted (column 2 of Fig. 12), but the road is not as homogeneous as the road extracted by automated IP-NMF (row 3, column 2 of Fig. 12). The  $3^{rd}$  extracted class (column 3 of Fig. 12) contains vegetation and shadowed asphalt. As in the previous case (automated IP-NMF) abundance maps show a combination of 2 subclasses in a class. With this manual initialisation, an improvement of the result is expected when the number of classes is increased, because the quality of automated IP-NMF results

falls due to the poor initialisation given by VCA. In Figure 14, seven classes of spectra were extracted: tile, asphalt, shadowed asphalt, tree, grass, shadowed tree, path (resp. columns 1 to 7 of Fig. 14). Compared to the automated IP-NMF results, the extraction of some classes is better performed here. Tiles are well retrieved despite their high variability, as the asphalt. However the vegetation extraction is better with the automated IP-NMF, indeed the grass ( $5^{th}$  column) and tree ( $4^{th}$  column) classes compete, which leads to poor tree extraction here. Besides the shadowed trees and shadowed tiles are extracted in the same class ( $7^{th}$  column) due to their spectral proximity.

So, it seems that a manual initialisation of the spectra with spectra belonging to expected classes improves the extraction of classes with high intra-class variability (tiles). However when intra-class variability is smaller, this initialisation can lead to lower performance than the automated one. For instance shadowed trees have a small inertia compared with the tiles, so the shadowed tree class can grow and other shadowed spectra are absorbed. However these spectra, even if they belong to the same class, still are spectrally different.

### 4) Result of automated IP-NMF with a post-processing:

Even with a good initialisation some problems persist. We propose to solve them by keeping the above automated IP-NMF method, followed by a post-processing stage. We exploit

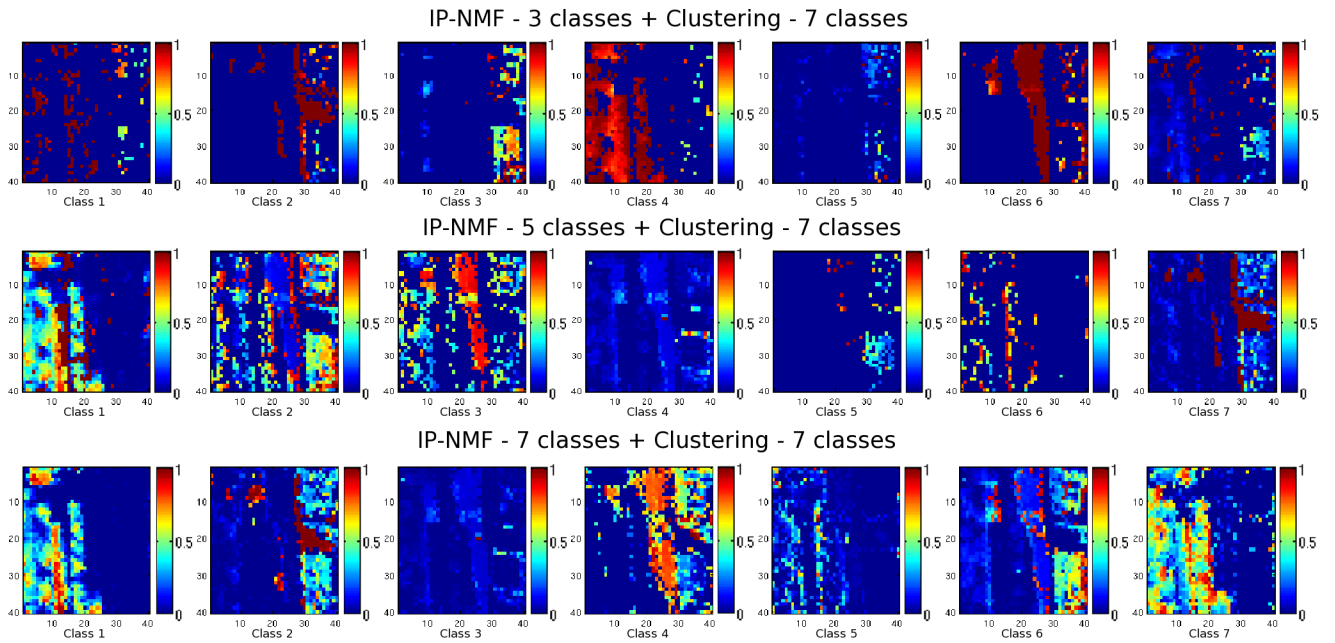


Fig. 15. Abundance maps obtained with automated IP-NMF applied with  $M = 3$  (1<sup>st</sup> line),  $M = 5$  (2<sup>nd</sup> line) and  $M = 7$  (3<sup>rd</sup> line) and followed by a clustering with 7 classes.

the fact that each pixel is described by a unique set of endmembers. After applying IP-NMF, we gathered all these spectra and then clustered them with the k-means method [31] using spectral angle as the similarity measure. The resulting abundance maps are depicted in Figure 15.

This post processing appears to be all the more efficient as the number of endmembers searched by IP-NMF is small. Indeed modifications of the maps between Fig. 14, 3<sup>rd</sup> line and Fig. 15, 3<sup>rd</sup> line (with classes extracted in arbitrary orders) are small compared to the differences between Fig. 12, 3<sup>rd</sup> line and Fig. 15, 1<sup>st</sup> line. This post-processing allows us splitting the classes composed of two subclasses. For instance, in the  $M = 3$  case, tile class and asphalt class are well separated. This post processing therefore allows one to further exploit the large amount of information extracted by IP-NMF, by improving the classification of the extracted pure spectra. Other post-processing methods can be imagined for other applications which would exploit the high number of spectra extracted with IP-NMF.

## VIII. CONCLUSION

In this paper, intra-class variability was first studied in order to characterise its magnitude. This analysis led us to develop a new mixing model taking this phenomenon into account. Two blind unmixing methods were then introduced to deal with intra-class variability. Tests on semi-synthetic data showed that UP-NMF has lower performance than IP-NMF. IP-NMF performance was also assessed with a real image. To process hyperspectral images, IP-NMF has the advantage of being automated, only the number of classes needs to be fixed. When initialisation of this method is provided by a usual unmixing method, IP-NMF always yields better performance than the latter method in our tests. Indeed

IP-NMF brings its flexibility to find endmembers around the single spectrum per class used to initialise it. The sensitivity of IP-NMF to the initial number of classes is lower than that of the considered classical unmixing methods. Besides, IP-NMF is able to find classes which were not detected by the above classical methods. This makes IP-NMF more robust to processing real images. Besides, to increase the accuracy of the results and to exploit the large number of endmembers provided by IP-NMF a post-processing stage can be added. Moreover, if users of this method agree to act on the algorithm, a manual initialisation could also be used instead of the above automated initialisation. This kind of initialisation allows the users to choose their classes of interest.

Our method is an extension of NMF. We chose to use a gradient descent to obtain our estimated spectra and abundances. Variants of this method may be developed by using other optimisation algorithms. The cost function optimised in this version penalises the spread of classes. Other versions could be imagined which would penalise the cost function with other terms (distance to the initial spectrum, introduction of spatial constraints ...). We plan to develop such versions and to compare them with the UP-NMF and IP-NMF methods.

## REFERENCES

- [1] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [2] R. B. Singer and T. B. McCord, "Mars - large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Lunar and Planetary Science Conference Proceedings*, vol. 10, 1979, pp. 1835–1848.
- [3] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear quadratic mixing model for reflectances in urban environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 544–558, Jan. 2014.

- [4] J. F. Mustard and C. M. Pieters, "Quantitative abundance estimates from bidirectional reflectance measurements," *J. Geophys. Res.*, vol. 92, no. B4, pp. E617–E626, Mar. 1987.
- [5] 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, no. 2, pp. 354–379, Apr. 2012.
- [6] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, Jan. 2014.
- [7] D. A. Roberts, "Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1712–1723, 2009.
- [8] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.
- [9] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.
- [10] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning Discriminative Sparse Representations for Modeling, Source Separation, and Mapping of Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4263–4281, Nov. 2011.
- [11] B. Somers, M. Zortea, A. Plaza, and G. 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, no. 2, pp. 396–408, Apr. 2012.
- [12] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [13] P. E. Dennison, K. Q. Halligan, and D. A. Roberts, "A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper," *Remote Sensing of Environment*, vol. 93, no. 3, pp. 359–367, Nov. 2004.
- [14] S. Lacherade, C. Miesch, X. Briottet, and H. L. Men, "Spectral variability and bidirectional reflectance behaviour of urban materials at a 20 cm spatial resolution in the visible and nearinfrared wavelengths. A case study over Toulouse (France)," *International Journal of Remote Sensing*, vol. 26, no. 17, pp. 3859–3866, Sep. 2005.
- [15] M. Veganzones, L. Drumetz, G. Tochon, M. Dalla Mura, A. Plaza, J. Bioucas-Dias, and J. Chanussot, "A new extended linear mixing model to address spectral variability," in *2014 6th Workshop on Hyperspectral Image and Signal Processing (WHISPERS)*, Lausanne, Switzerland, Jun. 2014.
- [16] 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, no. 4, pp. 898–910, Apr. 2005.
- [17] C. Bateson, G. Asner, and C. 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, Mar. 2000.
- [18] F. J. Garca-Haro, S. Sommer, and T. Kemper, "A new tool for variable multiple endmember spectral mixture analysis (VMESMA)," *International Journal of Remote Sensing*, vol. 26, no. 10, pp. 2135–2162, May 2005.
- [19] K. Adeline, X. Briottet, N. Paparoditis, and J.-P. Gastellu-Etchegorry, "Material reflectance retrieval in urban tree shadows with physics-based empirical atmospheric correction," in *Urban Remote Sensing Event (JURSE)*, Apr. 2013, pp. 279–283.
- [20] C. Miesch, L. Poutier, V. Achard, X. Lenot, and Y. Boucher, "Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1552–1562, Jul. 2005.
- [21] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.
- [22] P. O. Hoyer and P. Dayan, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [23] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007. [Online]. Available: <http://dx.doi.org/10.1162/neco.2007.19.10.2756>
- [24] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Tech. Rep., nov 2012, version 20121115.
- [25] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," vol. 3753, 1999, pp. 266–275.
- [26] 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, no. 3, pp. 529–545, Mar. 2001.
- [27] L. Zhang, B. Du, and Y. Zhong, "Hybrid detectors based on selective endmembers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2633–2646, Jun. 2010.
- [28] J. Broadwater and R. Chellappa, "Hybrid detectors for subpixel targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1891–1903, Nov. 2007.
- [29] S. Grupo de Inteligencia Computacional, Universidad del Pas Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), "Endmember induction algorithms (eias) toolbox."
- [30] J. Bioucas-Dias and J. Nascimento, "Hyperspectral Subspace Identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [31] J. MacQueen, "Some methods for classification and analysis of multivariate observations." The Regents of the University of California, 1967.