An NMF-based method for jointly handling mixture nonlinearity and intraclass variability in hyperspectral blind source separation

Yannick Deville[a,1,*], Guillaume Faury[a], Véronique Achard[b], Xavier Briottet[b]

[a]*Université de Toulouse, UPS, CNRS, CNES, Institut de Recherche en Astrophysique et Planétologie (IRAP), 14 Av. Edouard Belin, Toulouse, 31400, France*
[b]*ONERA DOTA, Université de Toulouse, 2 Av. Edouard Belin, Toulouse, 31400, France*

Abstract

Considering a set of observed signals that result from mixing (i.e. combining) a set of unknown source signals by means of an unknown function, blind source separation (BSS) and blind mixture identification (BMI) methods aim at estimating these source signals and/or the mixing function. Such methods have been extensively used, e.g. to process hyperspectral data in the field of remote sensing (i.e. Earth observation). However, most of these methods are restricted to the simplest mixing model, namely linear mixtures without the "variability" defined below. Some investigations started to tackle the more complex case of linear-quadratic (LQ), i.e. second-order polynomial, and memoryless mixtures. Some other investigations deal with the so-called intraclass variability phenomenon, i.e. with the case when each source yields a component having a somewhat different "shape" in each observed signal. Such nonlinearity and variability may e.g. appear in remote sensing hyperspectral images, as detailed in this paper. To our knowledge, each BSS/BMI method from the literature related to the nonlinearity and variability issues only addresses *one* of them. In contrast, we here propose a more powerful, very generic, BSS/BMI approach, that extends standard Nonnegative Matrix Factorization so as to *jointly* handle LQ mixtures and arbitrary source intraclass variability. This opens the way to a much wider range of BSS/BMI

applications than plain linear "variabilityless" mixtures, e.g. to handle the above-defined two issues of remote sensing. Considering the latter application, we built data sets so that they are both realistic and suited to a detailed quantitative performance evaluation. The tests reported in this paper show that our method significantly outperforms several approaches from the literature: the estimation errors are thus decreased by factors up to around 2 and 1.5, respectively for spectra and mixing coefficients.

1. Introduction

Blind source separation (BSS) and blind mixture identification (BMI) are mature fields within signal/image/data processing (see e.g. the books or survey chapters [8, 10, 12, 21, 32, 44]). They aim at solving problems where a set of unknown source signals get mixed, i.e. combined, through an unknown mixing function, thus yielding a set of known mixed signals, also called observations. One then aims at estimating the source signals and/or mixing function only from the known mixed signals, by also using some assumed properties about the source signals and/or mixing function.

The above type of problem is faced in various application fields, including remote sensing / Earth observation, when analyzing hyperspectral data (see e.g. the surveys in [6, 26]), for the following reason. Remote sensing hyperspectral data are obtained with sensors that have a low spatial resolution. Therefore, each pixel of such an image usually corresponds to a surface on Earth which is covered by several pure materials. The overall reflectance spectrum obtained for such a pixel is then a combination of the spectra of all involved pure materials. More precisely, the simplest configuration corresponds to homogeneous illumination of a flat surface on Earth, with sunlight directly reflected from the ground to the airbone or satellite sensor (and no intimate mixing, as discussed below) [13, 39]. Each pixel spectrum of an observed image is then a *linear* combination of the pure material spectra, with each associated coefficient equal to the fraction of surface covered by the considered pure material in this pixel. These coefficients are therefore called abundance fractions. A detailed analysis of hyperspectral images then requires one to derive two types of quantities from each recorded mixed pixel

spectrum: 1) the corresponding pure material spectra, that may then be further analyzed (classified) so as to identify the types of pure materials present in each pixel and 2) the corresponding abundance fractions, which may then e.g. be used to build a spatial map of the abundance of each pure material, that may be further analyzed, e.g. in order to manage crop in natural scenes or to derive the overall surface of solar panels in urban areas [25]. To this end, the pure material spectra and abundance fractions may be estimated by using suitable classes of BSS/BMI methods, often referred to as unsupervised unmixing methods by the remote sensing community. For the above-defined type of mixtures, a natural approach considers the pure spectra as the source signals, whereas the abundance fractions are the mixing coefficients. Alternatively, one may rewrite the data model so that the source signals are composed of the abundance fractions (with image pixels rearranged as a one-dimensional series) whereas the mixing coefficients then consist of the values of pure material spectra (the correspondence between these two models is e.g. detailed in [24]).

The above approach however has limited applicability because practical data often result from more complex mixing phenomena. Some configurations involve mixing at a microscopic level, referred to as intimate mixtures [26]. Even when such microscopic effects can be ignored (observations at a decimetric or metric scale), two other major phenomena may occur at the macroscopic level. The first one may still be considered in the framework of the most general BSS/BMI problem defined above. It concerns the case when the mixing function is *nonlinear*. Hyperspectral data give rise to such mixtures, and especially to linear-quadratic (LQ) ones, as shown e.g. in [35]. Briefly, this occurs when part of the sunlight is successively reflected by two surfaces before it reaches the sensor, e.g. in urban environments where it may be first reflected by the walls or windows of a building, down to the ground, and then from the ground to the sensor (see Fig. 1). In such conditions, [35] showed that the recorded pixel spectra are LQ mixtures (possibly restricted to bilinear ones, as defined in [16]) of the pure material spectra. Various methods have been proposed in the literature to solve the LQ version of the BSS/BMI problem in general (see e.g. the surveys in [11, 16]), and more specifically for unmixing hyperspectral data (see e.g. [1, 2, 3, 4, 14, 15, 17, 18, 20, 22, 23, 33, 34, 35, 36, 43, 45]).

Also starting from the standard linear mixing model, the other major macroscopic phenomenon that requires one to move to a more complex mixing model in various remote sensing configurations is the so-called intraclass
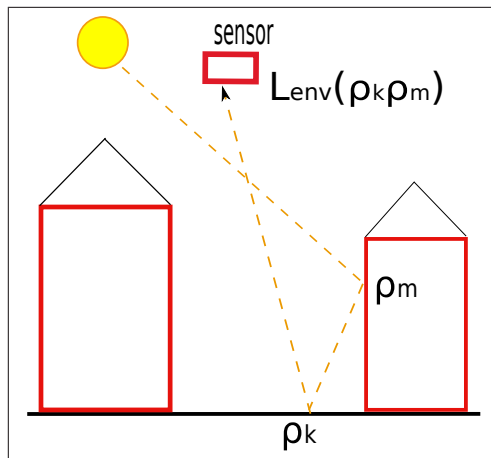
Figure 1: Double reflection of light emitted by the sun: the light is first reflected by a wall or window of a building, then reflected by the ground, and it eventually reaches the sensing device (Reprinted from [35]).

variability or spectral variability. In the above description, it was implicitly assumed that a given type of pure material is defined by a single spectrum so that, if the complete considered image involves M pure materials, then all observed pixel spectra are mixtures of the same M pure material spectra. The general definition of the BSS/BMI problem provided at the beginning of this section is also based on this assumption, the pure material spectra being then replaced by the source signals. However, things are more difficult in many practical remote sensing configurations, because what we called a pure material above yields somewhat different spectra in all observations, i.e. in all pixels: for instance, if various pixels contain roof tiles, the associated tile spectra are not the same for all these pixels (see e.g. [13, 39] and Section 4.1 below). One should then not think anymore in terms of a pure material defined by a single spectrum, but in terms of a class of materials (e.g. the class of roof tiles), whose spectra are somewhat different but still generally closer to one another than to the spectra of other classes of materials (e.g. the asphalt class if again considering urban areas). In other words, the classes have no or limited overlap (see an example e.g. in Fig. 5 of [39]). Starting from an image involving M classes of pure materials and containing P pixels, one may therefore aim at estimating MP pure spectra (that is, representatives of all classes in each pixel) and associated abundance fractions. Various methods have been proposed to this end: see e.g. the surveys [7, 48] and

4

recent original contributions e.g. in [30, 41, 42, 47]. Besides, [13, 39] describe a linear mixing model which includes a specific representative in each pixel for each class of pure materials, and a blind unmixing algorithm that handles variability by constraining the estimates of these representatives to have a limited spread within each class.

The previous work in [13, 39] is of importance in the framework of the present paper, because we here aim at extending that approach of [13, 39] so as to handle mixing nonlinearity in addition to variability. Indeed, when trying to extend unsupervised hyperspectral unmixing methods beyond the standard case of linear mixtures without variability, the papers cited above only considered one of the two possible issues, i.e. nonlinear mixtures or variability, but not both. In contrast, in the present paper, we propose an approach where we first define a data model which explicitly takes both phenomena into account and we then build an algorithm which allows one to estimate all the unknown variables involved in that model. This approach is based on one of the major classes of methods that have been developed in the literature to solve the BSS/BMI problem in general, and more specifically to perform unsupervised hyperspectral unmixing, namely Nonnegative Matrix Factorization (NMF) [5, 9, 10, 28, 29] (see also its origin in [38]). In particular, we here build upon the extension of NMF to LQ mixtures without variability that we proposed in [36]. The approach proposed in the present paper is thus very generic and could therefore be applied to other domains, in addition to Earth observation.

The remainder of this paper is organized as follows. In Section 2, we define the proposed data model. In Section 3, we introduce a corresponding unmixing method. Test results are reported in Section 4 and conclusions are drawn from this investigation in Section 5.

2. Data model

In this paper, the reflectance spectrum $\mathbf{x_p}$ associated with all L spectral bands and any pixel p of the observed image is modeled as the L-element

column vector

$$\mathbf{x_p} = \sum_{m=1}^{M} c_{p,m} \mathbf{r_m}(p) + \sum_{m=1}^{M} \sum_{\mu=m}^{M} c_{p,(m,\mu)} \mathbf{r_m}(p) \odot \mathbf{r}_\mu(p),$$

$$\forall\, p \in \{1,\dots,P\},$$

$$\text{with} \begin{cases} \sum_{m=1}^{M} c_{p,m} = 1 \\ c_{p,m} \geq 0 & \forall m \in \{1,\dots,M\} \\ 0 \leq c_{p,(m,\mu)} \leq 0.5 & \forall m, \mu, \text{ with } 1 \leq m \leq \mu \leq M \end{cases}$$

$$(1)$$

where the constraints on $c_{p,m}$ and $c_{p,(m,\mu)}$ are inspired from [35, 36], and with the following notations. M and P are, respectively, the number of classes of pure materials and the number of pixels in the observed image. $\mathbf{r}_m(p)$ is the pure material spectrum associated with the m-th class of materials and with pixel p. For a given class m, the corresponding pure material spectrum may depend on the considered pixel, thus allowing intraclass variability. Moreover, the model in itself allows any values for the spectra, which makes this model able to handle any type of variability (whereas some variability models from the literature are much more restricted, thus e.g. only allowing them to handle a global scale factor inside each class of pure materials: see e.g. [46]). The variability of the estimated model will be further constrained by the algorithm proposed in this paper. Besides, the proposed model (1) first includes a linear combination (with coefficients $c_{p,m}$) of the above spectra $\mathbf{r}_m(p)$, but also a linear combination (with coefficients $c_{p,(m,\mu)}$) of their element-wise products $\mathbf{r_m}(p) \odot \mathbf{r}_\mu(p)$. This allows this model to jointly include variability and quadratic effects (including bilinear ones, that correspond to $m \neq \mu$).

Still distinguishing between the linear and quadratic contributions, this yields a first matrix-form model for the complete observed data matrix $\mathbf{X} \in \mathbb{R}^{P \times L}$ composed of the vectors $\mathbf{x_p}$ for all pixels, rearranged as

$$\mathbf{X} = [\mathbf{x_1}, \dots, \mathbf{x_P}]^T \qquad (2)$$

where ^{T} stands for transpose. The proposed model then reads

$$\mathbf{X} = \tilde{\mathbf{C}}\tilde{\mathbf{R}} + \tilde{\boldsymbol{\Gamma}}\tilde{\mathbf{T}} \qquad (3)$$

with the following notations:

6

- $\mathbf{R}(p) = [\mathbf{r_1}(p), \dots, \mathbf{r_M}(p)]^T \in \mathbb{R}^{M \times L}$ contains all M pure material spectra for pixel p, with the spectrum for any class m stored in row m of $\mathbf{R}(p)$ whatever p.

- The matrix

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R}(1) \\ \vdots \\ \mathbf{R}(P) \end{bmatrix} \in \mathbb{R}^{PM \times L} \tag{4}$$

contains the pure material spectra for all pixels of the observed image, with the spectra for any class m stored in rows $m + (\kappa - 1)M$, with $\kappa \in \{1, \dots, P\}$.

- $\mathbf{c_p} = [c_{p,1}, \dots c_{p,m}]^T \in \mathbb{R}^{M \times 1}$ is the vector of linear mixing coefficients associated with $\mathbf{R}(p)$.

- $\tilde{\mathbf{C}} \in \mathbb{R}^{P \times PM}$ is the resulting block-diagonal matrix of linear mixing coefficients associated with $\tilde{\mathbf{R}}$ (i.e. with the complete image), defined as

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{c_1}^T & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & \mathbf{c_2}^T & \dots & 0 \dots 0 \\ & & \ddots & \\ 0 \dots 0 & 0 \dots 0 & \dots & \mathbf{c_P}^T \end{bmatrix}.$$

- $\mathbf{T}(p) = [\mathbf{r_1}(p) \odot \mathbf{r_1}(p), \mathbf{r_1}(p) \odot \mathbf{r_2}(p), \dots, \mathbf{r_M}(p) \odot \mathbf{r_M}(p)]^T \in \mathbb{R}^{K \times L}$, with $K = M^2 - \binom{M}{2}$, contains all element-wise products of pure material spectra (involved in quadratic components) for pixel p, in the same order for all pixels.

- The matrix

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{T}(1) \\ \vdots \\ \mathbf{T}(P) \end{bmatrix} \in \mathbb{R}^{PK \times L} \tag{5}$$

contains the element-wise products of pure material spectra for all pixels of the observed image.

- $\gamma_{\mathbf{p}} = [c_{p,(1,1)}, c_{p,(1,2)}, \dots, c_{p,(M,M)}]^T \in \mathbb{R}^{K \times 1}$ is the vector of quadratic mixing coefficients associated with $\mathbf{T}(p)$.

- $\tilde{\mathbf{\Gamma}} \in \mathbb{R}^{P \times PK}$ is the resulting block-diagonal matrix of quadratic mixing coefficients associated with $\tilde{\mathbf{T}}$ (i.e. with the complete image), defined as

$$\tilde{\mathbf{\Gamma}} = \begin{bmatrix} \gamma_{\mathbf{1}}^T & 0 \ldots 0 & \ldots & 0 \ldots 0 \\ 0 \ldots 0 & \gamma_{\mathbf{2}}^T & \ldots & 0 \ldots 0 \\ & & \ddots & \\ 0 \ldots 0 & 0 \ldots 0 & \ldots & \gamma_{\mathbf{P}}^T \end{bmatrix}.$$

Finally, this data model may be expressed in a more compact form, as

$$\mathbf{X} = \dot{\mathbf{C}}\dot{\mathbf{R}} \tag{6}$$

with the following notations:

- The matrix

$$\dot{\mathbf{R}} = \begin{bmatrix} \mathbf{R}(1) \\ \mathbf{T}(1) \\ \vdots \\ \mathbf{R}(P) \\ \mathbf{T}(P) \end{bmatrix} \in \mathbb{R}^{P(M+K) \times L} \tag{7}$$

contains the "extended spectra", i.e. the pure material spectra and their element-wise products, for all the observed image.

- Besides, $\dot{\mathbf{C}} \in \mathbb{R}^{P \times P(M+K)}$ is the block-diagonal matrix of linear and quadratic mixing coefficients associated with $\dot{\mathbf{R}}$, defined as

$$\dot{\mathbf{C}} = \begin{bmatrix} [\mathbf{c_1}^T, \gamma_{\mathbf{1}}^T] & 0 \ldots 0 & \ldots & 0 \ldots 0 \\ 0 \ldots 0 & [\mathbf{c_2}^T, \gamma_{\mathbf{2}}^T] & \ldots & 0 \ldots 0 \\ & & \ddots & \\ 0 \ldots 0 & 0 \ldots 0 & \ldots & [\mathbf{c_P}^T, \gamma_{\mathbf{P}}^T] \end{bmatrix}. \tag{8}$$

3. Proposed unmixing method

3.1. Principle of the method

We here extend the principles of linear NMF without variability, so as to derive a general BSS/BMI/unmixing method that jointly handles two phenomena, namely linear-quadratic (LQ) mixtures and intraclass variability. The latter phenomenon is dealt with by introducing an Inertia-constrained

Pixel-by-pixel (IP) approach, as detailed below, so that the proposed algorithm is called LQIP-NMF. This method involves two adaptive matrices that respectively aim at estimating the matrices $\dot{\mathbf{R}}$ and $\dot{\mathbf{C}}$ of the mixing model (6). These adaptive matrices might be denoted as $\widehat{\dot{\mathbf{R}}}$ and $\widehat{\dot{\mathbf{C}}}$ for the sake of clarity but, for better readability, they are simply denoted as $\dot{\mathbf{R}}$ and $\dot{\mathbf{C}}$ hereafter: we stess that, in all this Section 3, the notations $\dot{\mathbf{R}}$ and $\dot{\mathbf{C}}$ (and those for the associated variables, as defined in Section 2) refer to the *adaptive* matrices that are updated by the proposed unmixing algorithm, whereas the actual, *fixed*, values of the matrices in the right-hand term of the data model (6) are unknown and not used in this Section 3.

The LQIP-NMF method consists of adapting $\dot{\mathbf{R}}$ and $\dot{\mathbf{C}}$ so as to minimize the cost function defined as

$$J_{lqip-nmf}(\dot{\mathbf{R}}, \dot{\mathbf{C}}) = J_{RE}(\dot{\mathbf{R}}, \dot{\mathbf{C}}) + w J_I(\dot{\mathbf{R}}) \tag{9}$$

with

$$J_{RE}(\dot{\mathbf{R}}, \dot{\mathbf{C}}) = \frac{1}{2} \left\| \mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}} \right\|_F^2 \tag{10}$$

$$J_I(\dot{\mathbf{R}}) = \sum_{m=1}^{M} Tr(Cov(\tilde{\mathbf{R}}_{\mathcal{C}_{\mathbf{m}}})) \tag{11}$$

where $||.||_F$ stands for the Frobenius norm, Tr and Cov stand for matrix trace and covariance, and $\tilde{\mathbf{R}}_{\mathcal{C}_{\mathbf{m}}} \in \mathbb{R}^{P \times L}$ is the matrix containing all pure spectra estimates for the m-th class of materials and all pixels (the cost function (9) is minimized so as to take the constraints of model (1) into account, as detailed in the pseudo-code provided further in this paper). As shown by (10), $J_{RE}(\dot{\mathbf{R}}, \dot{\mathbf{C}})$ measures the reconstruction error achieved when approximating the observed image $\mathbf{X}$ by the product $\dot{\mathbf{C}}\dot{\mathbf{R}}$ of adaptive matrices. It extends the traditional NMF cost function to the much more complicated data structure defined by (7) and (8), which takes nonlinearity and variability into account. Besides, the penalty term $J_I(\dot{\mathbf{R}})$ defined in (11) is an extension of its restricted version only intended for *linear* mixtures with intraclass variability that we proposed in [39]. It is used to control the variability, inside each class of pure materials, allowed for the estimates of the pure spectra for that class that are associated with all pixels: as shown by (11), the quantity $J_I(\dot{\mathbf{R}})$ is the sum, over all classes, of the inertia (i.e. of the spread) $Tr(Cov(\tilde{\mathbf{R}}_{\mathcal{C}_{\mathbf{m}}}))$ of all estimated pure spectra for the considered class. Minimizing the overall

9

cost function (9) therefore also tends to force all the spectra estimated for one class to remain close to one another, and the strength of this constraint may be freely increased by increasing the value of the weight w in (9).

One should then define an optimization algorithm for minimizing the cost function $J_{lqip-nmf}(\dot{\mathbf{R}}, \dot{\mathbf{C}})$. A projected gradient algorithm is derived hereafter and the corresponding gradient expressions are therefore first calculated.

As with standard NMF, the proposed cost function may have spurious global minima and local minima. In order to make the proposed algorithm more likely to converge to an adequate point, one should initialize the adaptive matrices with relevant values. A practical initialization method is detail in Section 4 for the considered application.

3.2. Gradient with respect to $\tilde{\mathbf{R}}$: scalar form

To first calculate the partial derivative of the term J_{RE} of the cost function (9) with respect to $\tilde{\mathbf{R}}$, that term is rewritten as follows (see the proof in Appendix A):

$$J_{RE} = \sum_{p=1}^{P} J_p \quad \text{with } J_p = \frac{1}{2} \|\mathbf{x_p}^T - \begin{bmatrix} \mathbf{c_p}^T & \gamma_{\mathbf{p}}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix} \|_F^2. \tag{12}$$

This first yields the following scalar partial derivative of J_{RE} with respect to one element of $\mathbf{R}(p)$, with indices m and l:

$$\frac{\partial J_{RE}}{\partial \left[\mathbf{R}(p) \right]_{ml}} = \sum_{q=1}^{P} \frac{\partial J_q}{\partial \left[\mathbf{R}(p) \right]_{ml}}. \tag{13}$$

Moreover, (12) shows that J_q depends on $\mathbf{R}(p)$ only if $q = p$, so that (13) reduces to

$$\frac{\partial J_{RE}}{\partial \left[\mathbf{R}(p) \right]_{ml}} = \frac{\partial J_p}{\partial \left[\mathbf{R}(p) \right]_{ml}}. \tag{14}$$

It may then be shown that

$$\frac{\partial J_p}{\partial \left[\mathbf{R}(p)\right]_{ml}} = -[\mathbf{c_p}(\mathbf{x_p}^T - \left[\mathbf{c_p}^T \ \gamma_{\mathbf{p}}^T\right] \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix})]_{ml}$$

$$- \sum_{\substack{\mu=1 \\ \mu \neq m}}^{M} \left[\mathbf{R}(p)\right]_{\mu l} \times [\gamma_{\mathbf{p}}(\mathbf{x_p}^T - \left[\mathbf{c_p}^T \ \gamma_{\mathbf{p}}^T\right] \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix})]_{(\mu m)l}$$

$$- 2\left[\mathbf{R}(p)\right]_{ml} \times [\gamma_{\mathbf{p}}(\mathbf{x_p}^T - \left[\mathbf{c_p}^T \ \gamma_{\mathbf{p}}^T\right] \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix})]_{(mm)l}. \quad (15)$$

The above matrix $\mathbf{R}(p)$ is included in the matrix $\tilde{\mathbf{R}}$ defined by (4), and one then derives the scalar partial derivative of J_{RE} with respect to one element of $\tilde{\mathbf{R}}$, with indices ν and l. This may be shown to yield

$$\frac{\partial J_{RE}}{\partial [\tilde{\mathbf{R}}]_{\nu l}} = -\left[\tilde{\mathbf{C}}^T \left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{\nu l} - \sum_{\substack{\mu=(p-1)M+1 \\ \mu \neq \nu}}^{pM} \left[\tilde{\mathbf{R}}\right]_{\mu l} \times \left[\tilde{\mathbf{\Gamma}}^T \left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{(\mu\nu)l}$$

$$- 2\left[\tilde{\mathbf{R}}\right]_{\nu l} \times \left[\tilde{\mathbf{\Gamma}}^T \left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{(\nu\nu)l}$$

$$(16)$$

where (p, m) is the only couple of indices such that $\nu = (p - 1)M + m$, for a given ν with $\nu \in \{1, \ldots, PM\}$. In this couple (p, m), the index p is the index of the pixel with which $[\tilde{\mathbf{R}}]_{\nu l}$ is associated and m is the index of the class considered in $\mathbf{R}(p)$.

To then calculate the partial derivative of the term J_I of the cost function (9) with respect to $\tilde{\mathbf{R}}$, that term is rewritten as follows (see the proof and the expression of $Q_{\mathcal{C}_m}$ in Appendix B):

$$J_I = \frac{1}{P} Tr(\tilde{\mathbf{R}}^T \tilde{\mathbf{R}}) - \frac{1}{P^2} \sum_{m=1}^{M} Tr(Q_{\mathcal{C}_m}). \quad (17)$$

From this expression, tedious calculations show that

$$\frac{\partial J_I}{\partial \tilde{\mathbf{R}}} = \frac{2}{P}(\mathbf{Id}_{PM} - \frac{1}{P}\mathbf{U})\tilde{\mathbf{R}} \quad (18)$$

11

where $\mathbf{U} \in \mathbb{R}^{PM \times PM}$ is defined as

$$
\mathbf{U} = \begin{array}{c} \\ M\left\{ \phantom{\begin{matrix}1\\0\\\vdots\\0\\1\\\vdots\end{matrix}}\right. \end{array} \left[\begin{matrix} \overbrace{1 \quad 0 \quad \dots \quad 0}^{M} \quad 1 \quad \dots \\ 0 \quad 1 \quad \dots \quad 0 \quad 0 \quad \dots \\ \vdots \quad\quad \ddots \quad\quad \vdots \\ 0 \quad 0 \quad \dots \quad 1 \quad 0 \quad \dots \\ 1 \quad 0 \quad \dots \quad 0 \quad 1 \quad \dots \\ \vdots \quad\quad\quad\quad\quad \ddots \end{matrix}\right]
$$

$$
= \begin{bmatrix} \mathbf{Id}_M & \dots & \mathbf{Id}_M \\ \vdots & \ddots & \vdots \\ \mathbf{Id}_M & \dots & \mathbf{Id}_M \end{bmatrix} \tag{19}
$$

and $\mathbf{Id}_D$ is the identity matrix with dimension D. By using (16) and (18), Eq. (9) yields

$$
\left[\frac{\partial J_{lqip-nmf}}{\partial \tilde{\mathbf{R}}}\right]_{\nu l} = -\left[\tilde{\mathbf{C}}^T\left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{\nu l} - \sum_{\substack{\mu=(p-1)M+1 \\ \mu \neq \nu}}^{pM} \left[\tilde{\mathbf{R}}\right]_{\mu l} \times \left[\tilde{\mathbf{\Gamma}}^T\left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{(\mu\nu)l}
$$

$$
-2\left[\tilde{\mathbf{R}}\right]_{\nu l} \times \left[\tilde{\mathbf{\Gamma}}^T\left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right]_{(\nu\nu)l} + \frac{2w}{P}\left[\left(\mathbf{Id}_{PM} - \frac{1}{P}\mathbf{U}\right)\tilde{\mathbf{R}}\right]_{\nu l}. \tag{20}
$$

3.3. Gradient with respect to $\tilde{\mathbf{R}}$: matrix form

To take advantage of programming environments, such as Matlab, that use matrix representations to speed up computation, one must then use the scalar derivative (20) to derive the derivative of the cost function with respect to the complete matrix $\tilde{\mathbf{R}}$. Using the structure of $\tilde{\mathbf{R}}$ defined in (4), tedious calculations show that (20) yields

$$
\frac{\partial J_{lqip-nmf}}{\partial \tilde{\mathbf{R}}} = -\tilde{\mathbf{C}}^T\left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right) + \frac{2w}{P}\left(\mathbf{Id}_{PM} - \frac{1}{P}\mathbf{U}\right)\tilde{\mathbf{R}}
$$

$$
- \left(\mathbf{Id}_{PM} \otimes 1_{1K}\right)\left(\left(\left(\mathbf{Id}_P \otimes \mathbf{F}\right)\left(\mathbf{Id}_P \otimes \left(1_{M1} \otimes \mathbf{Id}_M\right)\right)\tilde{\mathbf{R}}\right)\right.
$$

$$
\left. \odot \left(\left(\mathbf{Id}_P \otimes \left(1_{M1} \otimes \mathbf{Id}_K\right)\right)\tilde{\mathbf{\Gamma}}\left(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}}\right)\right)\right)
$$

$$
\tag{21}
$$

12

where $\otimes$ is the Kronecker multiplication operator. The matrices denoted as 1_{YZ} contain $Y \times Z$ elements, all equal to 1. Besides, $\mathbf{F} \in \mathbb{R}^{MK \times M^2}$ is a block-diagonal matrix defined by

$$\mathbf{F} = \begin{bmatrix} \mathbf{F_1} & \cdots & [\mathbf{0}] \\ \vdots & \ddots & \vdots \\ [\mathbf{0}] & \cdots & \mathbf{F_M} \end{bmatrix}$$

$$[\mathbf{F}_\mu]_{km} = \sum_{\nu=1}^{M} \delta_m^\nu \delta_k^\eta (1 + \delta_\nu^\mu)$$

$$\eta = (min(\mu, \nu) - 1)(M - \frac{min(\mu, \nu)}{2}) + max(\mu, \nu)$$

where δ_i^j is the Kronecker symbol.

3.4. Update of matrix $\tilde{\mathbf{T}}$

The above equations will be used below to derive a gradient-based algorithm for updating the part of the matrix $\dot{\mathbf{R}}$, defined by (7), that corresponds to the matrix $\tilde{\mathbf{R}}$ defined by (4), i.e. that corresponds to all pure material spectra. In contrast, the part of $\dot{\mathbf{R}}$ that corresponds to the matrix $\tilde{\mathbf{T}}$ defined by (5) is not updated with a gradient-based rule, but in such a way that this part of $\dot{\mathbf{R}}$ contains all element-wise products of estimated pure material spectra. First using a scalar data representation, this yields the assignment

$$\left[\tilde{\mathbf{T}}^{(i+1)} \right]_{(m\nu)l} \leftarrow \left[\tilde{\mathbf{R}}^{(i+1)} \right]_{ml} \times \left[\tilde{\mathbf{R}}^{(i+1)} \right]_{\nu l} \tag{22}$$

where the superscript $^{(i+1)}$ refers to the new value of a matrix derived during iteration i of the unmixing algorithm.

Here again, the corresponding matrix form of this update rule is then preferred. Using the structure of $\tilde{\mathbf{T}}$ defined in (5), it may be shown to read

$$\tilde{\mathbf{T}} \longleftarrow \left((\mathbf{Id}_P \otimes \mathbf{G_1}) \tilde{\mathbf{R}} \right) \odot \left((\mathbf{Id}_P \otimes \mathbf{G_2}) \tilde{\mathbf{R}} \right) \tag{23}$$

where $\mathbf{G_1} \in \mathbb{R}^{K \times M}$ and $\mathbf{G_2} \in \mathbb{R}^{K \times M}$ are matrices defined as

$$[\mathbf{G_1}]_{km} = \sum_{\alpha=1}^{M} \sum_{\beta \geqslant \alpha}^{M} \delta_{(\alpha-1)(M-\frac{\alpha}{2})+\beta}^k . \delta_m^\alpha$$

$$[\mathbf{G_2}]_{km} = \sum_{\alpha=1}^{M} \sum_{\beta \geqslant \alpha}^{M} \delta_{(\alpha-1)(M-\frac{\alpha}{2})+\beta}^k . \delta_m^\beta .$$

3.5. Gradient with respect to $\dot{\mathbf{C}}$

Using matrix derivation properties, it is easily shown that

$$\frac{\partial J_{lqip-nmf}}{\partial \dot{\mathbf{C}}} = -(\mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}})\dot{\mathbf{R}}^T \qquad (24)$$

except that part of the elements of $\dot{\mathbf{C}}$ must be kept to zero, as detailed below.

3.6. Complete LQIP-NMF algorithm

The above calculations make it possible to derive the corresponding complete LQIP-NMF algorithm, whose pseudo-code is provided at the end of this section. The adaptive variables $\tilde{\mathbf{R}}$, $\tilde{\mathbf{T}}$ and $\dot{\mathbf{C}}$ are first initialized as detailed in Section 4 and then iteratively updated, denoting with superscripts $^{(i)}$ and $^{(i+1)}$ their values before and after adaptation during iteration i of that unmixing algorithm. At each iteration, $\tilde{\mathbf{R}}$ is first updated with the standard gradient algorithm (using (21) and an adaptation gain denoted as $\alpha_{\tilde{\mathbf{R}}}$) and each of its components is then projected onto the subset of positive values (by taking the maximum between that value and a very small positive real number ϵ). Then, $\tilde{\mathbf{T}}$ is updated with (23). The last steps of the algorithm aim at adapting $\dot{\mathbf{C}}$. They first use the standard gradient algorithm (and hence (24) and an adaptation gain denoted as $\alpha_{\dot{\mathbf{C}}}$), except that all "off-block-diagonal elements", i.e. all elements of $\dot{\mathbf{C}}^{(i+1)}$ that correspond to zero-valued elements of the data model (8), are then reset to zero. Finally, the elements of $\dot{\mathbf{C}}^{(i+1)}$ are post-processed so as to meet the inequalities and sum-to-one constraint of the data model (1), as detailed in Steps 4 and 5 of the pseudo-code shown below.

Algorithm: Linear-Quadratic Inertia-constrained Pixel-by-pixel NMF (LQIP-NMF)

$i \longleftarrow 1$
While ($J_{lqip-nmf} > Threshold$
and $i < Maximum\_number\_of\_iterations$):

1. Update matrix $\tilde{\mathbf{R}}$:

$$\tilde{\mathbf{R}}^{(i+1)} \longleftarrow \tilde{\mathbf{R}}^{(i)} + \alpha_{\dot{\mathbf{R}}} \left(\tilde{\mathbf{C}}^{(i)T} \left(\mathbf{X} - \dot{\mathbf{C}}^{(i)} \dot{\mathbf{R}}^{(i)} \right) \right.$$

$$- \frac{2w}{P} \left(\mathbf{Id}_{PM} - \frac{1}{P} \mathbf{U} \right) \tilde{\mathbf{R}}^{(i)} + (\mathbf{Id}_{PM} \otimes 1_{1K}) \times$$

$$\left(\left((\mathbf{Id}_P \otimes \mathbf{F}) (\mathbf{Id}_P \otimes (1_{M1} \otimes \mathbf{Id}_M)) \tilde{\mathbf{R}}^{(i)} \right) \odot \right.$$

$$\left. \left. \left((\mathbf{Id}_P \otimes (1_{M1} \otimes \mathbf{Id}_K)) \tilde{\mathbf{\Gamma}}^{(i)} \left(\mathbf{X} - \dot{\mathbf{C}}^{(i)} \dot{\mathbf{R}}^{(i)} \right) \right) \right) \right)$$

$$\tilde{\mathbf{R}}^{(i+1)} \longleftarrow max(\tilde{\mathbf{R}}^{(i+1)}, \epsilon)$$

2. Update matrix $\tilde{\mathbf{T}}$:
$$\tilde{\mathbf{T}}^{(i+1)} \longleftarrow \left((\mathbf{Id}_P \otimes \mathbf{G_1}) \tilde{\mathbf{R}}^{(i)} \right) \odot \left((\mathbf{Id}_P \otimes \mathbf{G_2}) \tilde{\mathbf{R}}^{(i)} \right)$$

3. Update matrix $\tilde{\mathbf{C}}$:
$$\dot{\mathbf{C}}^{(i+1)} \longleftarrow \dot{\mathbf{C}}^{(i)} + \alpha_{\dot{\mathbf{C}}} \left(\left(\mathbf{X} - \dot{\mathbf{C}}^{(i)} \dot{\mathbf{R}}^{(i)} \right) \dot{\mathbf{R}}^{(i)T} \right).$$
Then reset to zero all "off-block-diagonal elements" (see text).

4. Post-process linear coefficients:
For $p = 1$ to P
$$\mathbf{c_p}^{(i+1)} \longleftarrow max(\mathbf{c_p}^{(i+1)}, \epsilon)$$
$$\mathbf{c_p}^{(i+1)} \longleftarrow \mathbf{c_p}^{(i+1)} / \sum_{m=1}^{M} c_{p,m}$$

5. Post-process quadratic coefficients:
For $p = 1$ to P
   Update the vectors corresponding to all m and ν as follows:
$$\mathbf{c_{p,(m,\nu)}}^{(i+1)} \longleftarrow max(\mathbf{c_{p,(m,\nu)}}^{(i+1)}, \epsilon)$$
$$\mathbf{c_{p,(m,\nu)}}^{(i+1)} \longleftarrow min(\mathbf{c_{p,(m,\nu)}}^{(i+1)}, 0.5)$$

6. $i \longleftarrow i + 1$

4. Test results

4.1. Considered data

We here aim at accurately, and hence numerically, evaluating the performance of the proposed method. To this end, we process a matrix $\mathbf{X}$ of mixed

spectra, we thus get the estimated pure spectra and linear mixing coefficients (that could then be used as explained in Section 1 in final applications), and we compare these estimates to the actual values of these parameters. Implementing that protocol for a matrix $\mathbf{X}$ of *real*, i.e. measured, mixed data is quite challenging even when only considering the standard mixing model from the literature (i.e. the linear model without variability), because knowing the above actual values (i.e. ground truth) requests one to know all pure spectra involved in the considered scene and the abundance of each pure material in each pixel, typically derived by evaluating the surface covered by this material. It is then infeasible to extend the above protocol to *real* data $\mathbf{X}$ that obey the much more complex data model (1) considered in the present paper because, due to source variability, for each class of pure materials, one would need all versions of the corresponding spectrum for all pixels. To avoid that problem, we hereafter use mixed data $\mathbf{X}$ that are synthetic but realistic, since they combine real pure spectra that have variability, by using the model (1).

More precisely, the reported tests involve $M = 3$ classes of pure materials, namely tiles, vegetation and asphalt, which is relevant for urban applications. For each class, a complete set of presumably pure spectra were manually extracted from a real urban hyperspectral image from the city of Toulouse (France). Most features of this image are the same as in [39] and the considered area is shown in Figure 2. The database of presumably pure spectra here consists of 190 tile spectra, 55 vegetation spectra and 52 asphalt spectra. This set of spectra yields large intraclass variabilities. The angular variability is here separately evaluated for each of the 3 classes, by means of the Spectral Angle Mapper (SAM) [27] with respect to the mean of the spectra of the considered class. The mean and maximum values of this SAM over a class are respectively equal to 2.64° and 8.53° for the tile class, 7.38° and 15.39° for the vegetation class, 2.73° and 10.05° for the asphalt class. In contrast, it is generally considered that spectra corresponding to the same class of materials yield SAMs limited to a few degrees. For instance, for default settings, the SAM-based classification method available in the commercial ENVI software [31] considers that spectra from a class have a SAM lower than 0.1rad $\simeq 5.73°$, with respect to the representative of that class. The larger values faced in our database may result from the fact that a few spectra that were manually selected (based on the objects present in the considered scene) are in fact not completely pure.

Hereafter, we will take advantage of that property to also analyze the per-

16

Figure 2: Image of the considered urban area (Toulouse, France) (Reprinted from [39]).

formance of the considered unmixing methods depending on the magnitude of variability, by also considering various subsets of this overall set of spectra, that have a lower variability. To this end, we select a value f, ranging from 0% to 100%, that defines the fraction of spectra that we want to keep, separately for each class of pure materials. The kept spectra are then chosen as follows. Each class c contains a total number of spectra denoted as $n(c)$. We only keep the spectra of that class which yield a SAM, with respect to the mean spectrum of that complete class, that is lower than a bound, where we tune this bound so that the number of kept spectra is almost equal to $f \times n(c)$. The mean and maximum SAM values, that were considered above only for each *complete* class, can now also be computed for each subset of data per class corresponding to a given value of f (now computing SAMs with respect to the mean of spectra over that fixed subset). Their variations with respect to the fraction f are provided in Fig. 3. This shows that, to obtain subsets of data that correspond to realistic classes, i.e. to SAM limited to a few degrees, f should be upper bounded by a value around 80% or even lower (depending on the considered class, depending whether one focuses on the mean or maximum SAM in the class and depending whether one sets the SAM limit at $\simeq 5.73°$ as mentioned above). The minimum, median and maximum spectra obtained for each class with $f = 80\%$ are shown in Fig. 4. Each subset of pure spectra selected with the above-defined approach is then used as follows. For each pixel, one pure spectrum is randomly selected from the considered subset for each class, to derive the corresponding mixed spectrum of the data matrix $\mathbf{X}$, which contains 756 pixels. The mixing coefficient

17

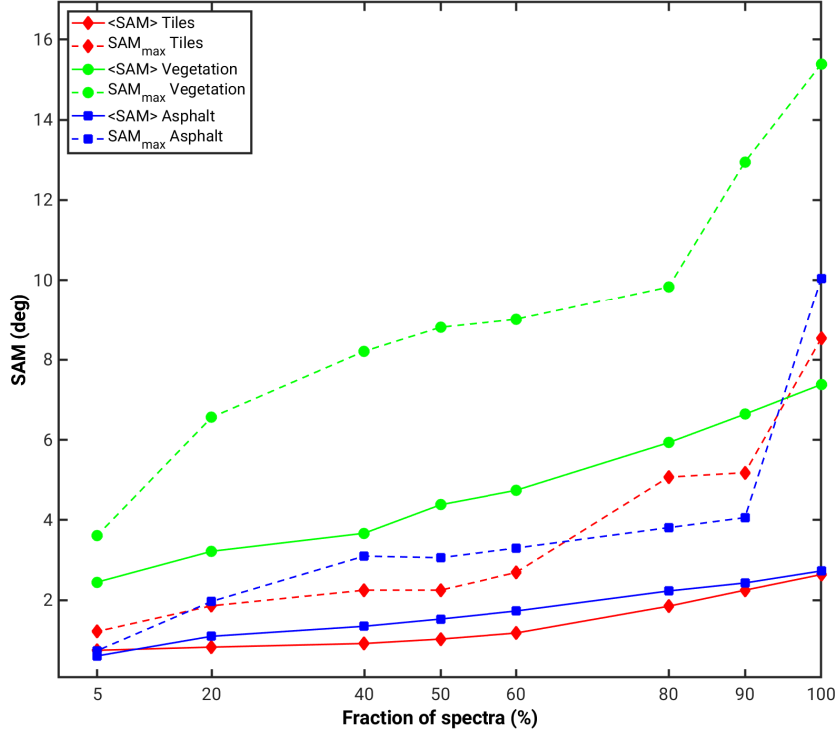values used to this end in (1) are defined as follows.

Figure 3: For each class of materials, mean ($<$SAM$>$) and maximum (SAM$_{\text{max}}$) of SAM (in degrees), versus fraction f of selected spectra (see text).

The LQIP-NMF method can handle the complete linear-quadratic mixing model, i.e. here with 3 linear and 6 quadratic estimated coefficients per pixel, and it is indeed operated in this way hereafter. However, as a simple synthetic case, the actual data matrix $\mathbf{X}$ considered here is restricted to the bilinear mixing model. This means that non-zero quadratic coefficients are used in $\mathbf{X}$ only for element-wise products of *different* pure spectra, that therefore involve the (tile, vegetation), (tile, asphalt) and (vegetation, asphalt) pairs of pure spectra. Moreover, some linear and quadratic coefficients in $\mathbf{X}$ are set to zero, especially to create a few pure pixels (i.e. with one linear coefficient equal to one, whereas all the other linear coefficients and all quadratic ones
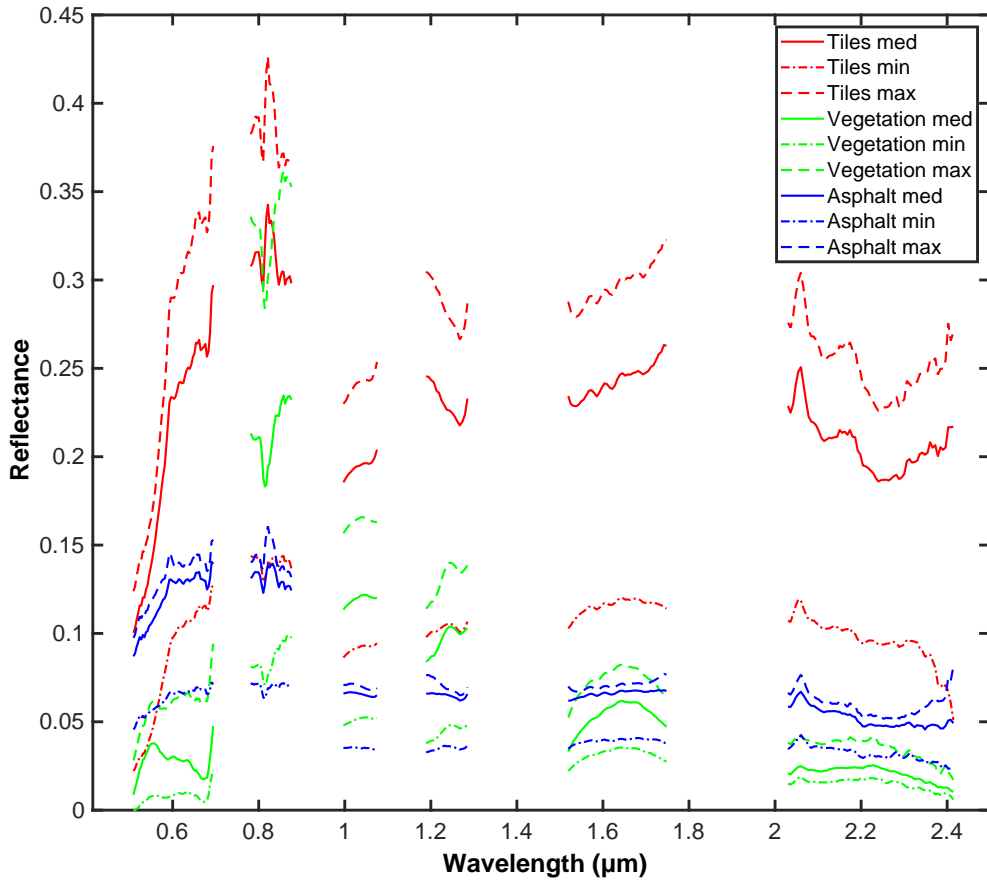
Figure 4: Minimum, median and maximum spectra obtained for each class with $f = 80\%$.

are equal to zero) and to ensure that, when a bilinear term has a non-zero coefficient in a pixel, the corresponding two linear terms also have non-zero coefficients in that pixel. In each non-pure pixel, the non-zero linear mixing coefficients are arbitrarily selected and scaled so that they sum to one. Besides, the non-zero quadratic coefficients are randomly drawn with a uniform distribution between 0 and an upper bound γ_{max}, that is varied in our tests in order to investigate the dependence of performance with respect to the degree of nonlinearity of the mixing model.

4.2. Tested methods

The above-defined data are first processed with the LQIP-NMF algorithm and a variant, both operated as follows. As in standard NMF methods, the matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{C}}$ used here need to be initialized before they are adapted. This is here performed with the following procedure. The standard VCA unmixing algorithm [37], intended for linear mixtures, is first applied to the matrix $\mathbf{X}$ of mixed data, with a number of estimated spectra set to $M = 3$, thus neglecting the nonlinear part actually present in the processed data matrix $\mathbf{X}$ at this stage. That part is then taken into account when setting the initial value of the matrix $\widehat{\mathbf{R}}$ of extended spectra: for each pixel, the $M = 3$ "linear rows of $\widehat{\mathbf{R}}$" (see the data structure in (7)) are set to the spectra estimated by VCA, whereas the $M(M + 1)/2 = 6$ "quadratic rows of $\widehat{\mathbf{R}}$" are set to the element-wise products of the spectra provided by VCA. All the linear mixing coefficients contained in the matrix $\widehat{\mathbf{C}}$ are initially set to $1/M$, so that they meet the standard sum-to-one constraint. The initial values of all estimated quadratic mixing coefficients in $\widehat{\dot{\mathbf{C}}}$ are randomly drawn (independently from the actual values in $\dot{\mathbf{C}}$), with a uniform distribution between 0 and the upper bound γ_{max}.

After the above initializations, two alternative methods are first considered for updating the matrices $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{C}}$. The first one is the complete LQIP-NMF algorithm defined in Section 3.6. The second one, called LQIP-NMF-FCLSU below, is a hybrid method, where the update rule of the LQIP-NMF algorithm is only used to update $\widehat{\mathbf{R}}$ whereas, at each iteration, the coefficient matrix $\widehat{\dot{\mathbf{C}}}$ is derived from $\mathbf{X}$ and from the current value of $\widehat{\mathbf{R}}$ as follows, separately considering each pixel p of the data matrix $\mathbf{X}$. The FCLSU algorithm [19] is first applied to the considered pixel p of $\mathbf{X}$ and to the part of the current matrix $\widehat{\mathbf{R}}$ corresponding to pixel p, with 9 abundance fraction coefficients summing to one for this pixel p, as if the considered data consisted of linear mixtures of the above-defined set of 9 extended spectra for pixel p. Then, the 6 abundance fractions corresponding to products of "standard spectra" are left unchanged, whereas the same scale factor is applied to the 3 abundance fractions corresponding to these "standard spectra" themselves, so that the resulting 3 coefficients sum to one. This protocol is used because the linear-quadratic model developed in [35] was shown to meet the sum-to-one condition for its linear abundance fractions, whereas quadratic

20

abundance fractions "come in addition".

Two types of algorithms derived from the literature were tested in addition to the above methods, namely:

1. An unmixing method which is intended neither for variability nor nonlinearity in the data. This method consists of a combination of VCA (for estimating spectra) and then FCLSU (for estimating abundance fractions), with a number of estimated spectra set to $M = 3$ and 3 estimated linear abundances per pixel, thus neglecting the nonlinear part actually present in the processed data matrix $\mathbf{X}$. This method is called VCA-FCLSU hereafter.

2. Two unmixing methods intended for variability but not for nonlinearity in the data. The first one is the IP-NMF algorithm defined in [13, 39], which here estimates 3 pure spectra and 3 linear coefficients that are all specific to each pixel. The second one, called IP-NMF-FCLSU below, is an original hybrid method, where the update rule of the IP-NMF algorithm is only used to update the matrix that contains 3 spectra per pixel whereas, at each iteration, the matrix that contains 3 linear coefficients per pixel is derived from $\mathbf{X}$ and from the current value of the above spectral matrix, by using the FCLSU algorithm separately for each pixel. Here again, the variables used to store estimated pure spectra are initialized with the results of VCA, whereas the variables used to store estimated mixing coefficients are initialized to $1/M$.

4.3. Performance criteria

The performance of the tested unmixing methods is evaluated by means of two criteria. First, the quality of the estimated pure spectra is measured by comparing them to the actual pure spectra used to create the mixed matrix $\mathbf{X}$. This comparison is performed by using the SAM. This SAM parameter is first averaged over all $M = 3$ spectra and all image pixels. Moreover, for each considered configuration, 100 runs are performed with different randomly drawn data in each run. The SAM values provided hereafter are averages derived from all these runs.

Besides, the quality of the estimated linear mixing coefficients is evaluated thanks to the (mean) Coefficient Error (CE), obtained by first computing the Frobenius norm of the difference between the vectors of actual and estimated coefficients, then dividing that norm by the number of pixels in the image, and finally again averaging these results over all 100 runs.

4.4. Results

In all the tests we performed, the LQIP-NMF-FCLSU and IP-NMF-FCLSU methods yielded better performance than their LQIP-NMF and IP-NMF variants. Therefore, we hereafter only consider LQIP-NMF-FCLSU and IP-NMF-FCLSU, together with the VCA-FCLSU method. Moreover, we present two types of tests: 1) tests performed in the noiseless case, i.e. with a data matrix $\mathbf{X}$ that obeys the model (6), which does not include noise, and then 2) tests with a noise component added in each pixel to the component $\dot{\mathbf{C}}\dot{\mathbf{R}}$ of (6) to form the final noisy data matrix $\mathbf{X}$.

4.4.1. Tests with noiseless data

Within the noiseless case, a first set of tests was performed with a fixed degree of nonlinearity in the observed data, whereas their spectral variability was varied. To this end, the parameter γ_{max} defined in Section 4.1 was set to 0.3. This corresponds to a rather high degree of nonlinearity, with a significant number of quadratic coefficients ranging up to 0.3. This may e.g. be compared with the investigation of urban scenes in [35]: although the upper bound of quadratic coefficients is then 0.5, their histogram is a decreasing function, with most coefficients concentrated on significantly lower values, up to 0.15 for most of them and 0.25 for some others. Besides, we here control the variability of these data by using the protocol defined in Section 4.1, keeping in mind that the relevant variability domain is obtained for f lower than or equal to around 80%. The values of the SAM and CE parameters thus obtained with the considered three methods are shown in Fig. 5 and 6. These figures show that the proposed LQIP-NMF-FCLSU method yields significantly better performance than the IP-NMF-FCLSU and VCA-FCLSU methods derived from the literature, whatever the variability in the relevant range defined at most by $f \in [0, 80\%]$ and even for higher values of f.

A complementary set of tests was then performed with a varying degree of nonlinearity in the observed data, whereas their spectral variability was fixed. To this end, γ_{max} was varied from 0.1 to 0.5, thus covering a wide range of nonlinearities. Meanwhile, f was fixed to 80%, which corresponds to a large variability, as explained above. The values of SAM and CE thus obtained with the considered three methods are shown in Fig. 7 and 8. LQIP-NMF-FCLSU here again significantly outperforms IP-NMF-FCLSU and VCA-FCLSU, whatever the considered degree of nonlinearity.
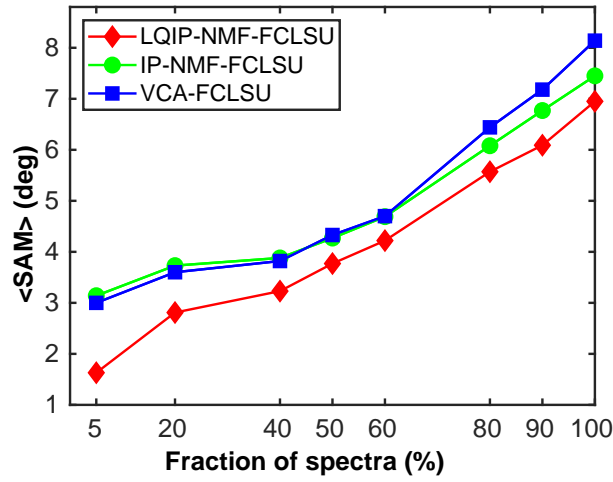
Figure 5: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean SAM (in degrees), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text).
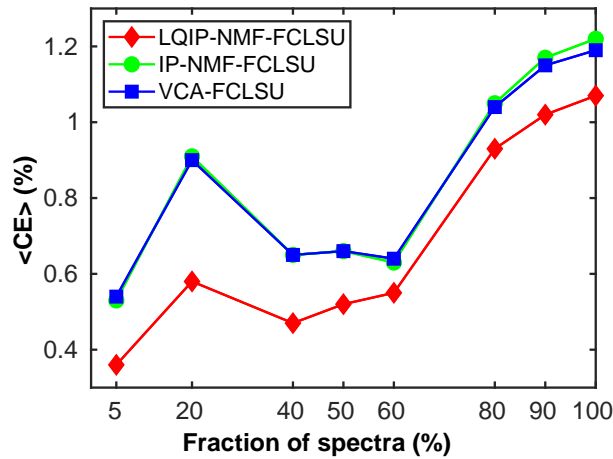

Figure 6: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean Coefficient Error (CE, in %), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text).

A more detailed analysis of the obtained results can be performed by graphically comparing the actual pure material spectra used to create the processed data matrix $\mathbf{X}$ with their estimates derived by the considered un-
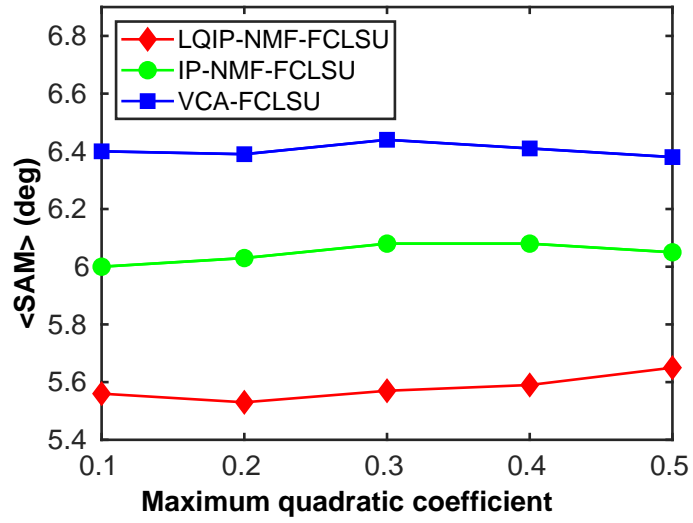
Figure 7: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean SAM (in degrees), versus upper bound γ_{max} of quadratic mixing coefficients, for $f = 80\%$.
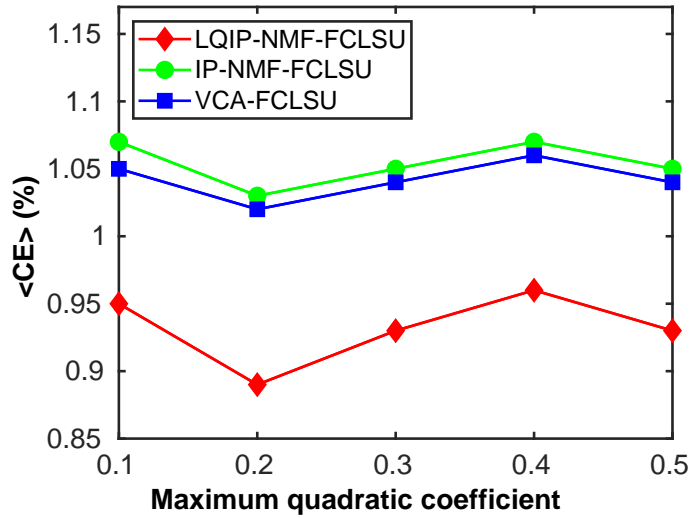

Figure 8: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean Coefficient Error (CE, in %), versus upper bound γ_{max} of quadratic mixing coefficients, for $f = 80\%$.

mixing method. Focusing on the proposed LQIP-NMF-FCLSU method, we provide such results in Fig. 9 to 11. These results are obtained in the difficult case involving both high spectral variability (with $f = 80\%$) and high nonlinearity (with $\gamma_{max} = 0.3$). Fig. 9 to 11 show average performance. To this end, each of these figures addresses only one of the considered classes of pure material spectra (tile, vegetation or asphalt) and shows the actual and estimated spectra in the pixel where the SAM between these two spectra takes the median value, among the values corresponding to all pixels of the image. When disregarding the scale factors, since they are not taken into account in the SAM parameter, Fig. 9 to 11 show that the considered actual and estimated spectra are quite similar. This confirms the good performance of the proposed LQIP-NMF-FCLSU method, even in the difficult considered case.
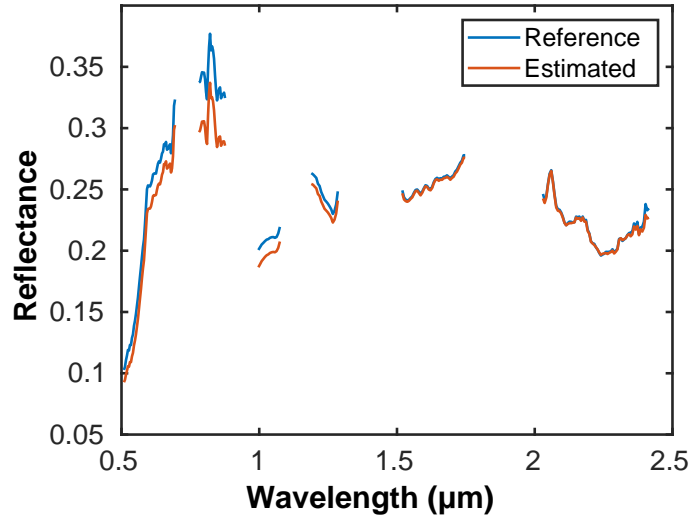

Figure 9: Actual spectrum and spectrum estimated by LQIP-NMF-FCLSU, for tile class, in the pixel which yields the median value of SAM for that class (SAM = 2.74°), for $f = 80\%$ and $\gamma_{max} = 0.3$.
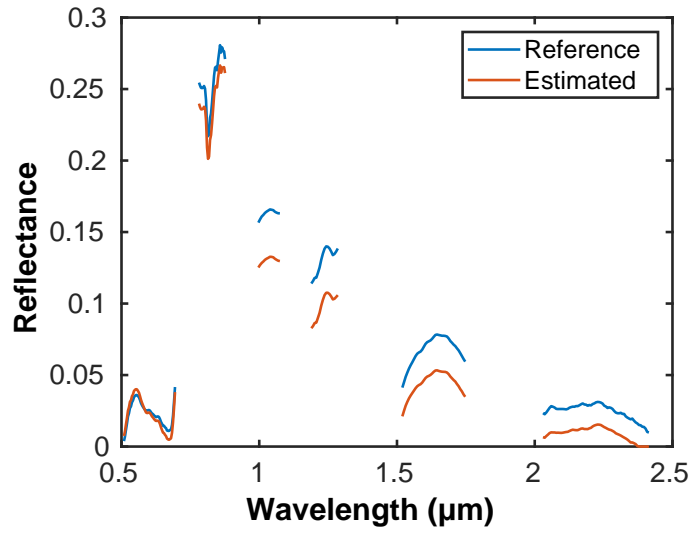
Figure 10: Actual spectrum and spectrum estimated by LQIP-NMF-FCLSU, for vegetation class, in the pixel which yields the median value of SAM for that class (SAM = 7.55°), for $f = 80\%$ and $\gamma_{max} = 0.3$.
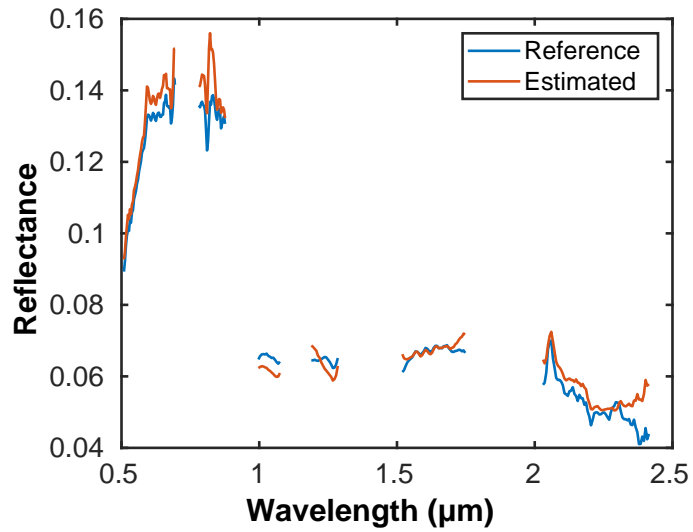

Figure 11: Actual spectrum and spectrum estimated by LQIP-NMF-FCLSU, for asphalt class, in the pixel which yields the median value of SAM for that class (SAM = 2.42°), for $f = 80\%$ and $\gamma_{max} = 0.3$.

   As explained above, we also performed tests with a noise component added in each pixel to the component $\dot{\mathbf{C}}\dot{\mathbf{R}}$ of (6) to form the final noisy data matrix $\mathbf{X}$. These noise components were zero mean and uniformly distributed over a range whose width was successively selected so as to achieve a signal-to-noise ratio (SNR) equal to 40 dB or 30 dB over all the image. These SNR values were selected because, in the studied application to hyperspectral remote sensing, it is generally considered that the SNR of real-world data may decrease down to around 40 dB, or possibly 30 dB. As in the first type of test reported in Section 4.4.1, the degree of nonlinearity is here fixed, with $\gamma_{max} = 0.3$, whereas the spectral variability is varied, by varying f. The results thus obtained are shown in Fig. 12 to 15. These figures, together with Fig. 5 and 6 for the noiseless case, show that the performance of all considered methods only slightly degrades when the SNR decreases in the considered range. Moreover, whatever the SNR, the proposed LQIP-NMF-FCLSU method significantly outperforms the other approaches.
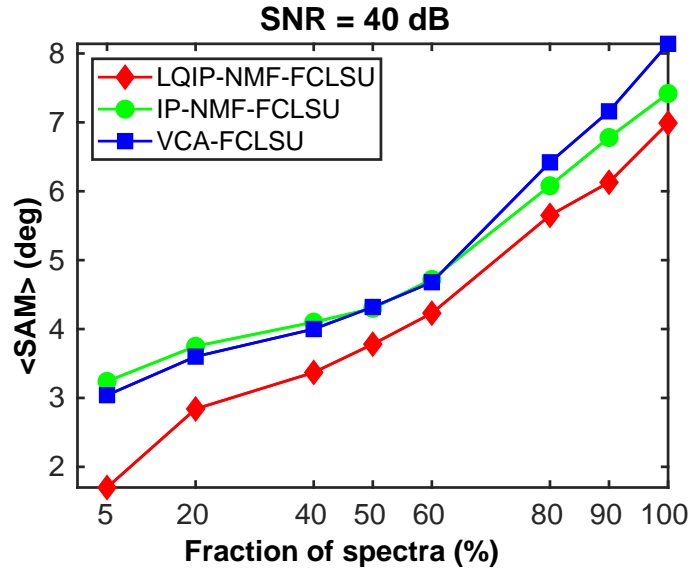

Figure 12: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean SAM (in degrees), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text) and Signal-to-Noise Ratio (SNR) = 40 dB.
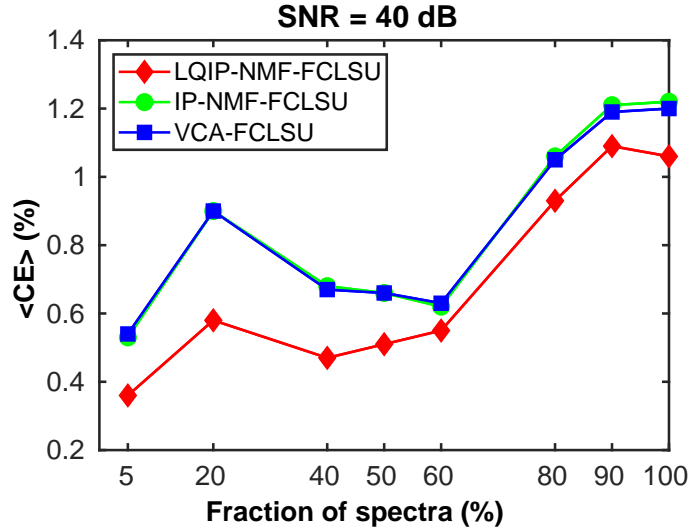
Figure 13: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean Coefficient Error (CE, in %), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text) and Signal-to-Noise Ratio (SNR) = 40 dB.

4.4.3. Computational complexity

The computational complexity of all tested methods was evaluated for a laptop including an 11th generation Intel Core i7-1185G7 (3.00 GHz x 8), with 31 GBytes of memory. The mean execution time per run (averaged over all 100 runs) was around 0.1 s, 2.7 s and 13.1 s, respectively for the VCA-FCLSU, IP-NMF-FCLSU and LQIP-NMF-FCLSU methods. As expected, the methods that address spectral variability, namely IP-NMF-FCLSU and LQIP-NMF-FCLSU, yield significantly higher computational times, due to the much larger sizes of some matrices that they use. Also as expected, to handle mixture non-linearity in addition, LQIP-NMF-FCLSU has a somewhat higher complexity than IP-NMF-FCLSU. Anyway, all methods result in quite limited execution times per run: at most a few seconds. Their computational load is therefore not a major issue, so that one can focus on the accuracy of the considered methods and one can thus reach the overall conclusion that the proposed LQIP-NMF-FCLSU method significantly outperforms the other approaches.
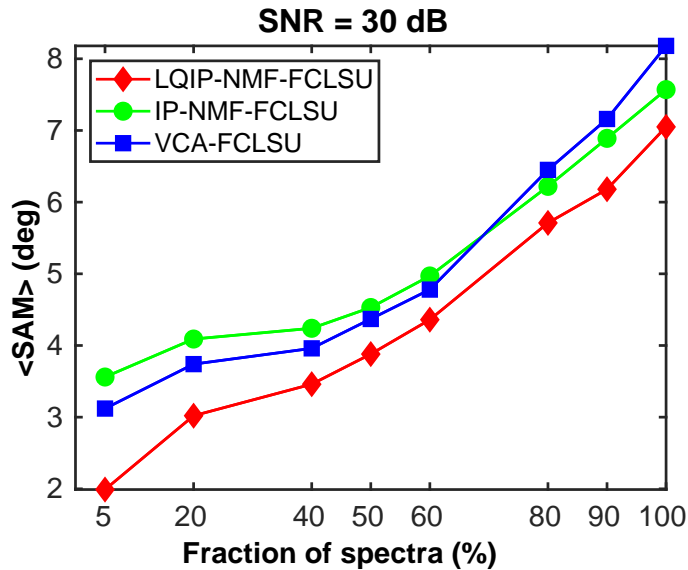
Figure 14: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean SAM (in degrees), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text) and Signal-to-Noise Ratio (SNR) = 30 dB.

5. Conclusion and future work

The Blind Source Separation (BSS) and Blind Mixture Identification (BMI) methods reported in the literature have especially been developed to handle the simplest mixing model, whereas we here proposed a very generic NMF-based method that extends previous works so as to *jointly* address two extensions of that model: (i) we moved from linear mixtures to second-order polynomial ones, called linear-quadratic mixtures, and (ii) we considered the case when the source signals have so-called intraclass variability, i.e. each source yields (here, first- and second-order) components having a somewhat different "shape" in each observed signal. Both phenomena are actually faced in practical applications, including unsupervised hyperspectral unmixing in Earth observation, which was one of our major motivations for developing such methods. We therefore validated the performance of the proposed method for hyperspectral data, showing that it significantly outperforms approaches from the literature. We plan to also test this method for a higher number of classes of pure materials.

The proposed framework may be further extended by considering higher-
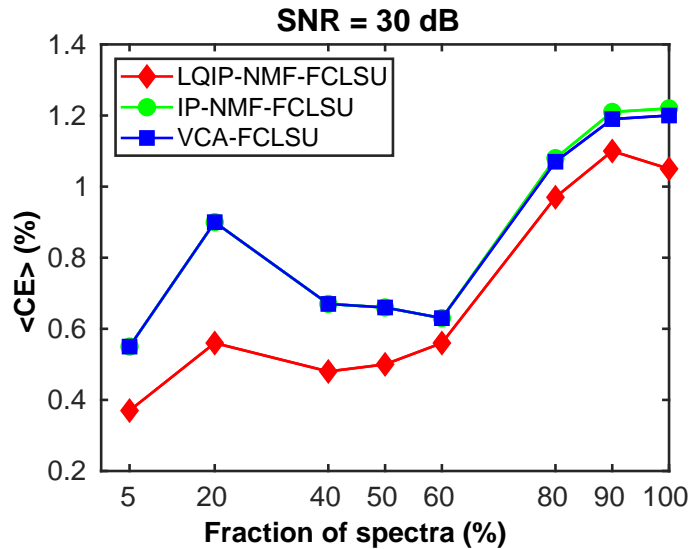
29

Figure 15: Performance of LQIP-NMF-FCLSU, IP-NMF-FCLSU and VCA-FCLSU methods: mean Coefficient Error (CE, in %), versus fraction f of selected spectra, for $\gamma_{max} = 0.3$ (see text) and Signal-to-Noise Ratio (SNR) = 30 dB.

order polynomial mixtures combined with intraclass variability. This may e.g. be useful in the framework of Earth observation, to handle multiple reflections of light, beyond the second-order reflections that only give rise to the second-order polynomial mixtures considered in the present paper (higher-order reflections however yield signal components that have lower powers). Some BSS/BMI investigations have been reported in the literature for higher-order polynomial mixtures, but without intraclass variability (see references in [16]). In particular, in [34] we showed how to extend NMF without variability to mixtures involving third-order polynomials. That investigation [34] is a good starting point for further extending that approach so as to handle intraclass variability in addition.

Appendix A. Other expression of J_{RE}

$$J_{RE} = \frac{1}{2} \left\| \mathbf{X} - \dot{\mathbf{C}}\dot{\mathbf{R}} \right\|_F^2$$

$$= \frac{1}{2} \left\| \begin{bmatrix} \mathbf{x_1}^T \\ \ldots \\ \mathbf{x_P}^T \end{bmatrix} - \begin{bmatrix} \mathbf{c_1}^T \mathbf{R}(1) + \gamma_1^T \mathbf{T}(1) \\ \ldots \\ \mathbf{c_P}^T \mathbf{R}(P) + \gamma_P^T \mathbf{T}(P) \end{bmatrix} \right\|_F^2$$

$$= \frac{1}{2} \sum_{\substack{1 \le p \le P \\ 1 \le l \le L}} \left(\begin{bmatrix} \mathbf{x_1}^T - \mathbf{c_1}^T \mathbf{R}(1) - \gamma_1^T \mathbf{T}(1) \\ \ldots \\ \mathbf{x_P}^T - \mathbf{c_P}^T \mathbf{R}(P) - \gamma_P^T \mathbf{T}(P) \end{bmatrix}_{pl} \right)^2$$

$$= \frac{1}{2} \sum_{p=1}^P \sum_{l=1}^L \left(\left[\mathbf{x_P}^T - \mathbf{c_P}^T \mathbf{R}(p) - \gamma_P^T \mathbf{T}(p) \right]_l \right)^2$$

$$= \frac{1}{2} \sum_{p=1}^P \left\| \mathbf{x_P}^T - \begin{bmatrix} \mathbf{c_P}^T & \gamma_P^T \end{bmatrix} \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix} \right\|_F^2$$

$$= \sum_{p=1}^P J_p \tag{A.1}$$

with

$$J_p = \frac{1}{2} \left\| \mathbf{x_P}^T - \begin{bmatrix} \mathbf{c_P}^T & \gamma_P^T \end{bmatrix} \begin{bmatrix} \mathbf{R}(p) \\ \mathbf{T}(p) \end{bmatrix} \right\|_F^2. \tag{A.2}$$

Appendix B. Other expression of J_I

$$J_I = \sum_{m=1}^M Tr(Cov(\tilde{\mathbf{R}}_{\mathcal{C}_\mathbf{m}}))$$

$$= \sum_{m=1}^M \left(\frac{1}{P} Tr(\tilde{\mathbf{R}}_{\mathcal{C}_\mathbf{m}}^T \tilde{\mathbf{R}}_{\mathcal{C}_\mathbf{m}}) - \frac{1}{P^2} Tr(Q_{\mathcal{C}_m}) \right) \tag{B.1}$$

$$Q_{\mathcal{C}_m} = \tilde{\mathbf{R}}_{\mathcal{C}_\mathbf{m}}^T 1_{P,P} \tilde{\mathbf{R}}_{\mathcal{C}_\mathbf{m}} \tag{B.2}$$

where $1_{P,P}$ is the $P \times P$ matrix with all elements equal to 1. Hence

$$
\begin{aligned}
J_I &= \sum_{m=1}^{M} \left(\frac{1}{P} \sum_{l=1}^{L} \sum_{k=1}^{P} [\tilde{R}_{\mathcal{C}_m}]_{k,l}^2 - \frac{1}{P^2} Tr(Q_{\mathcal{C}_m}) \right) \\
&= \frac{1}{P} \sum_{m=1}^{M} \sum_{l=1}^{L} \sum_{k=1}^{P} [\tilde{R}]_{(k-1)M+m,l}^2 - \frac{1}{P^2} \sum_{m=1}^{M} Tr(Q_{\mathcal{C}_m}) \\
&= \frac{1}{P} \sum_{\kappa=1}^{PM} \sum_{l=1}^{L} [\tilde{R}]_{\kappa,l}^2 - \frac{1}{P^2} \sum_{m=1}^{M} Tr(Q_{\mathcal{C}_m}) \\
&= \frac{1}{P} Tr(\tilde{\mathbf{R}}^T \tilde{\mathbf{R}}) - \frac{1}{P^2} \sum_{m=1}^{M} Tr(Q_{\mathcal{C}_m}).
\end{aligned}
\tag{B.3}
$$

References

[1] F. Z. Benhalouche, Y. Deville, M. S. Karoui, A. Ouamri, "Hyperspectral endmember spectra extraction based on constrained linear-quadratic matrix factorization using a projected gradient method", Proceedings of the 2016 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2016), pp. 1-6, Salerno, Italy, Sept. 13-16, 2016.

[2] F. Z. Benhalouche, M. S. Karoui, Y. Deville, "Linear-quadratic NMF-based urban hyperspectral data unmixing with some known endmembers", Proceedings of the 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2018), Amsterdam, The Netherlands, Sept. 23-26, 2018.

[3] F. Z. Benhalouche, Y. Deville, M. S. Karoui, A. Ouamri, "Hyperspectral unmixing based on constrained bilinear or linear-quadratic matrix factorization", Remote Sensing, vol. 13, issue 11, paper no. 2132, 2021. DOI: https://doi.org/10.3390/rs13112132

[4] Y. K. Benkouider, F. Z. Benhalouche, M. S. Karoui, Y. Deville, S. Hosseini, "Bilinear matrix factorization using a gradient method for unmixing hyperspectral images combined with multispectral data", the 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2018), Amsterdam, The Netherlands, Sept. 23-26, 2018.

[5] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization", Computational statistics & data analysis, vol. 52, no. 1, pp. 155-173, Sept. 2007.

[6] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, 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, 2012.

[7] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, C. Jutten, "Spectral variability in hyperspectral data unmixing. A comprehensive review", IEEE Geoscience and Remote Sensing Magazine, pp. 2-49, 2021.

[8] A. Cichocki, S.-I. Amari, "Adaptive blind signal and image processing. Learning algorithms and applications", Wiley, Chichester, England, 2002.

[9] A. Cichocki, R. Zdunek, A. H. Phan, S.-I. Amari, Nonnegative matrix and tensor factorizations. Applications to exploratory multi-way data analysis and blind source separation, Wiley, Chichester, UK, 2009.

[10] P. Comon and C. Jutten Eds, "Handbook of blind source separation. Independent component analysis and applications", Academic Press, Oxford, UK, 2010.

[11] Y. Deville, L. T. Duarte, "An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures", Proceedings of the 12th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2015), Springer International Publishing Switzerland, LNCS 9237, pp. 155-167, Liberec, Czech Republic, Aug. 25-28, 2015.

[12] Y. Deville, "Blind source separation and blind mixture identification methods", Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1-33, J. Webster (ed.), Wiley, 2016. DOI: 10.1002/047134608X.W8300

[13] Y. Deville, C. Revel, V. Achard, X. Briottet, "Application and extension of PCA concepts to blind unmixing of hyperspectral data with intra-class variability", chapter in "Advances in Principal Component Analysis - Research and Development", pp 225-252, G. R. Naik (ed), Springer, Singapore, 2018. DOI: https://doi.org/10.1007/978-981-10-6704-4_9

[14] Y. Deville, "From separability/identifiability properties of bilinear and linear-quadratic mixture matrix factorization to factorization algorithms", Digital Signal Processing, vol. 87, pp. 21-33, April 2019. DOI: https://doi.org/10.1016/j.dsp.2019.01.011

[15] Y. Deville, S. Hosseini, "Blind source separation methods based on output nonlinear correlation for bilinear mixtures of an arbitrary number of possibly correlated signals", Proceedings of the Eleventh IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2020), Hangzhou, China, June 8-11, 2020.

[16] Y. Deville, L. T. Duarte, S. Hosseini, "Nonlinear blind source separation and blind mixture identification. Methods for bilinear, linear-quadratic and polynomial mixtures", SpringerBriefs in Electrical and Computer Engineering, Springer Nature, 2021. DOI: 10.1007/978-3-030-64977-7

[17] O. Eches, M. Guillaume, "A bilinear-bilinear nonnegative matrix factorization method for hyperspectral unmixing", IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 4, pp. 778-782, April 2014.

[18] A. Guerrero, Y. Deville, S. Hosseini, "A blind source separation method based on output nonlinear correlation for bilinear mixtures", Proceedings of the 14th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2018), Springer International Publishing AG, part of Springer Nature 2018, LNCS 10891, pp. 183-192, Guildford, UK, July 2-5, 2018.

[19] D. Heinz, 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.

[20] P. Huard, R. Marion, "Study of non-linear mixing in hyperspectral imagery - a first attempt in the laboratory", Proceedings of the Third

Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2011), Lisbon, Portugal, June 6-9, 2011.

[21] A. Hyvärinen, J. Karhunen, E. Oja, "Independent Component Analysis", Wiley, New York, 2001.

[22] L. Jarboui, S. Hosseini, Y. Deville, R. Guidara, A. Ben Hamida, "A new unsupervised method for hyperspectral image unmixing using a linear-quadratic model", Proceedings of the First International Conference of Advanced Technologies for Signal and Image Processing (ATSIP 2014), pp. 423-428, Sousse, Tunisia, March 17-19, 2014.

[23] L. Jarboui, S. Hosseini, R. Guidara, Y. Deville, A. Ben Hamida, "A MAP-based NMF approach to hyperspectral image unmixing using a linear-quadratic mixture model", Proceedings of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), pp. 3356-3360, Shanghai, China, March, 20-25, 2016.

[24] M. S. Karoui, Y. Deville, S. Hosseini, A. Ouamri, "Blind spatial unmixing of multispectral images: New methods combining sparse component analysis, clustering and non-negativity constraints", Pattern Recognition, vol. 45, pp. 4263-4278, 2012.

[25] M. S. Karoui, F. Z. Benhalouche, Y. Deville, K. Djerriri, X. Briottet, T. Houet, A. Le Bris, C. Weber, "Partial linear NMF-based unmixing methods for detection and area estimation of photovoltaic panels in urban hyperspectral remote sensing data", Remote Sensing, Vol. 11, Issue 18, September-2, paper no. 2164, 2019. doi:10.3390/rs11182164

[26] N. Keshava, J.F. Mustard, "Spectral unmixing", IEEE Signal Processing Magazine, pp. 44-57, Jan. 2002.

[27] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, A. F. H. Goetz, "The spectral image processing system (SIPS) - Interactive visulaization and analysis of imaging spectrometer data", Remote Sensing of Environment, vol. 44, pp. 145-163, 1993.

[28] D. D. Lee, H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", Nature, vol. 401, pp. 788-791, 1999.

[29] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization", Adv. Neural Info. Proc. Syst. 13, pp. 556-562, 2001.

[30] H. Liu, Y. Lu, Z. Wu, Q. Du, J. Chanussot, Z. Wei, "Bayesian unmixing of hyperspectral image sequence with composite priors for abundance and endmember variability", to appear in the IEEE Transactions on Geoscience and Remote Sensing.

[31] L3HARRIS Geospatial, ENVI software,
https://www.l3harrisgeospatial.com/docs/spectralanglemapper.html

[32] S. Makino, T.-W. Lee, H. Sawada (Eds), "Blind speech separation", Springer, Dordrecht, The Netherlands, 2007.

[33] I. Meganem, P. Déliot, X. Briottet, Y. Deville, S. Hosseini, "Physical modelling and non-linear unmixing method for urban hyperspectral images", Proceedings of the Third Workshop on Hyperspectral Image and Signal Processing (WHISPERS 2011), Lisbon, Portugal, June 6-9, 2011.

[34] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, L. T. Duarte, "Linear-quadratic and polynomial Non-negative Matrix Factorization; application to spectral unmixing", Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011), Barcelona, Spain, Aug. 29 - Sept. 2, 2011.

[35] I. Meganem, P. Déliot, X. Briottet, Y. Deville, 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.

[36] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, "Linear-quadratic blind source separation Using NMF to unmix urban hyperspectral images", IEEE Transactions on Signal Processing, vol. 62, no. 7, pp. 1822-1833, April 1, 2014.

[37] J. M. P. Nascimento, J. M. 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, April 2005.

[38] P. Paatero, U. Tapper, P. Aalto, M. Kulmala, "Matrix factorization methods for analysing difussion battery data", J. Aerosol Sci., vol. 22, suppl. 1, pp. S273-S276, 1991.

[39] C. Revel, Y. Deville, V. Achard, X. Briottet, C. Weber, "Inertia-constrained pixel-by-pixel nonnegative matrix factorisation: a hyperspectral unmixing method dealing with intra-class variability", Remote Sensing, Vol. 10, Issue 11, 1706, Nov. 2018. DOI: https://doi.org/10.3390/rs10111706

[40] C. Revel, "Apport de la prise en compte de la variabilité intra-classe dans les méthodes de démélange hyperspectral pour l'imagerie urbaine", Thèse de l'Université Paul Sabatier de Toulouse, France, Dec. 19, 2016.

[41] Y. E. Salehani, E. Arabnejad, S. Gazor, "Augmented Gaussian linear mixture model for spectral variability in hyperspectral unmixing", Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021), pp. 1880-1884, Toronto, ON, Canada, June 6-11, 2021.

[42] S. Shi, M. Zhao, L. Zhang, J. Chen, "Variational autoencoders for hyperspectral unmixing with endmember variability", Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021), pp. 1875-1879, Toronto, ON, Canada, June 6-11, 2021.

[43] J. Sigurdsson, M. O. Ulfarsson, J. R. Sveinsson, "Blind nonlinear hyperspectral unmixing using an l_q regularizer", Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018), pp. 4229-4232, Valencia, Spain, July 22-27, 2018.

[44] A. Smilde, R. Bro, P. Geladi, "Multi-way analysis with applications in the chemical sciences", Wiley, Chichester, England, 2004.

[45] Y. Su, J. Li, H. Qi, P. Gamba, A. Plaza, J. Plaza, "Multi-task learning with low-rank matrix factorization for hyperspectral nonlinear unmixing", Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), pp. 2127-2130, Yokohama, Japan, July 28 - Aug. 2, 2019.

[46] M. A. Veganzones, L. Drumetz, G. Tochon, M. Dalla Mura, A. Plaza, J. Bioucas-Dias, J. Chanussot, "A new extended linear mixing model to address spectral variability", Proceedings of the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), pp. 1-4, Lausanne, Switzerland, June 24-27, 2014.

[47] J. Yu, B. Wang, Y. Lin, F. Li, J. Cai, "A novel inequality-constrained weighted linear mixture model for endmember variability", Remote Sensing of Environment vol. 257, 112359, 2021.

[48] A. Zare and K.C. Ho, "Endmember variability in hyperspectral analysis", IEEE Signal Processing Magazine, vol. 31, no. 1, pp. 95-104, Jan. 2014.