# Hyperspectral Unmixing with Spectral Variability Using a Perturbed Linear Mixing Model

Pierre-Antoine Thouvenin, *Student Member, IEEE*, Nicolas Dobigeon, *Senior Member, IEEE* and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—Given a mixed hyperspectral data set, linear unmixing aims at estimating the reference spectral signatures composing the data – referred to as *endmembers* – their abundance fractions and their number. In practice, the identified endmembers can vary spectrally within a given image and can thus be construed as variable instances of reference endmembers. Ignoring this variability induces estimation errors that are propagated into the unmixing procedure. To address this issue, endmember variability estimation consists of estimating the reference spectral signatures from which the estimated endmembers have been derived as well as their variability with respect to these references. This paper introduces a new linear mixing model that explicitly accounts for spatial and spectral endmember variabilities. The parameters of this model can be estimated using an optimization algorithm based on the alternating direction method of multipliers. The performance of the proposed unmixing method is evaluated on synthetic and real data. A comparison with state-of-the-art algorithms designed to model and estimate endmember variability allows the interest of the proposed unmixing solution to be appreciated.**

***Index Terms*—Hyperspectral imagery, linear unmixing, endmember spatial and spectral variability, *Alternating Direction Method of Multipliers* (ADMM).**

## I. Introduction

OVER the past decades, hyperspectral imagery has been receiving an increasing interest. Whereas traditional red / green / blue or multispectral images are composed of a limited number of spectral channels (from three to tens), hyperspectral images are acquired in hundreds of contiguous spectral bands facilitating the analysis of the elements in the scene, e.g., determining their nature and relative proportions. However, the high spectral resolution of these images is mitigated by their lower spatial resolution, hence the need to unmix the data. Spectral unmixing is aimed at estimating the reference spectral signatures – referred to as endmembers – their abundance fractions and their number from which the $L$-multi-band observations are derived according to a predefined mixing model. Assuming the absence of any microscopic interaction between the materials of the imaged scene and a negligible declivity, a linear mixing model (LMM) is classically used to describe the structure of the collected data [1]. However, the spectral signatures contained in a reference image can vary spectrally, spatially or temporally from an image to another due to varying acquisition conditions. This can result in significant estimation errors being propagated throughout the unmixing process. Various models either derived from a statistical or a deterministic point of view have been designed to address this issue [2]. More precisely, the first class of methods assumes that the endmember spectra can be considered as realizations of multivariate distributions. The most popular models are the normal composition model [3] and the beta compositional model [4]. The second class of methods considers the endmember signatures as members of spectral libraries associated with each material (bundles). Two methods using spectral libraries have been especially considered in the literature: the automated endmember bundles (AEB) [5] and the Fisher discriminant null space (FDNS) [6]. Whereas AEB enables the extraction of an endmember library to account for spectral variabilities, the aim of FDNS is to estimate a transformation projection matrix to project the hyperspectral data into a space minimizing the variability impact.

Since the identified endmembers can be considered as variable instances of reference endmembers, we introduce an extended version of the classical LMM to explicitly model the spectral variability. In [7], the variability is assumed to only result from scaling factors. Conversely, in this paper, inspired by a model designed in [8], each endmember is represented by a "pure" spectral signature corrupted by an additive perturbation accounting for its variability. The perturbation is allowed to vary from a pixel to another to represent spatial-spectral variabilities. As a result, the proposed perturbed LMM (PLMM) can capture endmember spatial and spectral variability within a given image. To the best of our knowledge, it is the first time endmember variability has been explicitly modeled as an additive perturbation.

The promising results obtained with the alternating direction method of multipliers (ADMM) in hyperspectral imagery [9] and in image deblurring [10]–[14] serve as an incentive to apply a similar framework to conduct PLMM-based unmixing. A key property of the ADMM framework lies in the introduction of appropriate splitting variables. Indeed, the specified constraints can be handled independently from the rest of the problem and often lead to analytical solutions when solving the resulting optimization problem. Using this fruitful principle, an ADMM-based algorithm for linear unmixing using a group lasso $\ell_{2,1}$-norm regularization was recently developed in [15], [16]. Inspired by these examples, this paper proposes to exploit the advantages of an ADMM-based resolution of the



This work is supported by the Direction Générale de l'Armement, French Ministry of Defence.

The authors are with the University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: {pierreantoine.thouvenin, Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr

TABLE I
TABLE OF NOTATIONS.

| | |
|---|---|
| $N$ | number of pixels |
| $L$ | number of spectral bands |
| $K$ | number of endmembers |
| $\mathbf{Y} \in \mathbb{R}^{L\times N}$ | lexicographically ordered pixels |
| $\mathbf{M} \in \mathbb{R}^{L\times K}$ | endmember matrix |
| $\mathbf{dM}_n \in \mathbb{R}^{L\times K}$ | $n$th variability matrix |
| $\mathbf{A} \in \mathbb{R}^{K\times N}$ | abundance matrix |
| $\mathbf{V} \in \mathbb{R}^{(K-1)\times L}$ | projector on the space spanned by the $K-1$ principal components of $\mathbf{Y}$ |
| $\mathbf{U} \in \mathbb{R}^{L\times(K-1)}$ | inverse projector |
| $\mathbf{T} \in \mathbb{R}^{(K-1)\times K}$ | projection of $\mathbf{M}$ on the PCA space |
| $\mathbf{m}_{n,k}$, $\widetilde{\mathbf{m}}_{n,j}$ | column $k$, row $j$ of the matrix $\mathbf{M}_n$ |
| $[\mathbf{AB}]_k$, $\{\mathbf{AB}\}_j$ | column $k$, row $j$ of the matrix $\mathbf{AB}$ |
| $\succeq$ | term-wise inequality |
| $\mathcal{S}^+_{n,p}$ | $= \{\mathbf{X} \in \mathbb{R}^{n\times p} \vert \mathbf{X} \succeq \mathbf{0}_{n,p}\}$ |
| $\mathbf{1}_n$ | $= [1, 1, \ldots, 1]^T \in \mathbb{R}^n$ |
| $\mathcal{I}_A(x)$ | $= \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise.} \end{cases}$ |

linear unmixing problem to account for spatial and spectral endmember variabilities.

Throughout the article, the number of endmembers will be assumed to be *a priori* known or estimated by any state-of-the-art method (e.g., [17]) since estimating the required number of endmembers to appropriately describe the data as well as endmember variability is a challenging task. Indeed, the choice of $K$ drastically alters the representation of the imaged scene, and is thus a crucial step to the endmember identification and the subsequent abundance estimation [1], [17], [18].

The paper is organized as follows. The PLMM accounting for spectral and spatial variabilities is introduced in Section II. Section III describes an ADMM-based algorithm to solve the resulting optimization problem. Experimental results obtained on synthetic and real data are reported in Section IV and V respectively. The results obtained with the proposed algorithm are systematically compared to those of the vertex component analysis / fully constrained least squares (VCA/FCLS), the simplex identification via split augmented Lagrangian (SISAL) [9] coupled with FCLS, AEB and FDNS. Section VI finally concludes this work.

## II. PROBLEM STATEMENT

### A. Perturbed linear mixing model (PLMM)

In the absence of any specific prior knowledge on the variability nature (i.e., errors affecting the endmembers), we have chosen to explicitly represent the variability by a spatially varying additive endmember perturbation. This choice, inspired by a model designed in [8], appears to be simple and flexible enough to account for the observed variability. Assuming that the number of endmembers $K$ is known, the proposed PLMM differs from the classical LMM insofar as each pixel $\mathbf{y}_n$ is represented by a combination of the $K$ endmembers – denoted as $\mathbf{m}_k$ – affected by a perturbation vector $\mathbf{dm}_{n,k}$ accounting for endmember variability. The resulting PLMM can be written

$$\mathbf{y}_n = \sum_{k=1}^{K} a_{kn}\Big(\mathbf{m}_k + \mathbf{dm}_{n,k}\Big) + \mathbf{b}_n \text{ for } n = 1, \ldots, N \quad (1)$$

where $\mathbf{y}_n$ denotes the $n$th image pixel, $\mathbf{m}_k$ is the $k$th endmember, $a_{kn}$ is the proportion of the $k$th endmember in the $n$th pixel, and $\mathbf{dm}_{n,k}$ denotes the perturbation of the $k$th endmember in the $n$th pixel. Finally, $\mathbf{b}_n$ models a zero-mean white Gaussian noise resulting from the data acquisition as well as modeling errors. We can note that the proposed PLMM reduces to the classical LMM in absence of variability. In matrix form, the PLMM (1) can be written as follows

$$\mathbf{Y} = \mathbf{MA} + \underbrace{\Big[\ \mathbf{dM}_1\mathbf{a}_1 \ \Big\vert\ \ldots \ \Big\vert\ \mathbf{dM}_N\mathbf{a}_N\ \Big]}_{\boldsymbol{\Delta}} + \mathbf{B} \quad (2)$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ is an $L \times N$ matrix containing the image pixels, $\mathbf{M}$ is an $L \times K$ matrix containing the endmembers, $\mathbf{A}$ is a $K \times N$ matrix composed of the abundance vectors $\mathbf{a}_n$, $\mathbf{dM}_n$ is an $L \times K$ matrix whose columns are the perturbation vectors associated with the $n$th pixel, and $\mathbf{B}$ is an $L \times N$ matrix accounting for the noise. The non-negativity and sum-to-one constraints usually considered to reflect physical considerations are

$$\begin{aligned} &\mathbf{A} \succeq \mathbf{0}_{K,N}, \quad \mathbf{A}^T\mathbf{1}_K = \mathbf{1}_N \\ &\mathbf{M} \succeq \mathbf{0}_{L,K}, \quad \mathbf{M} + \mathbf{dM}_n \succeq \mathbf{0}_{L,K},\ \forall n = 1, \ldots, N. \end{aligned} \quad (3)$$

When compared to the underlying models proposed in the literature to mitigate variability [2], model (1) presents the advantage to explicitly address the variability phenomenon in terms of an additive perturbation affecting each endmember. This perturbation accounts for any deviation from the linear mixing model (as will be illustrated in our experiments). The main contribution of this paper is to propose an unsupervised algorithm for estimating the endmembers contained in the image and the abundances and endmember variability for each pixel of this image.

### B. Problem formulation

As mentioned in Section I, the PLMM (1) and constraints (3) can be combined to formulate a constrained optimization problem. An appropriate cost function is required to estimate the parameters $\mathbf{M}, \mathbf{A}, \mathbf{dM}$. Assuming the signal is corrupted by a zero-mean white Gaussian noise, we define the data fitting term as the Frobenius norm of the difference between the acquisitions $\mathbf{Y}$ and the reconstructed data $\mathbf{MA} + \boldsymbol{\Delta}$. Since the problem is ill-posed, additional penalization terms are needed. In this paper, we propose to define penalization functions $\Phi, \Psi$ and $\Upsilon$ to reflect the available *a priori* knowledge on $\mathbf{M}, \mathbf{A}$ and $\mathbf{dM}$ respectively. As a result, the optimization problem is expressed as

$$(\mathbf{M}^*, \mathbf{dM}^*, \mathbf{A}^*) \in \underset{\mathbf{M},\mathbf{dM},\mathbf{A}}{\arg\min} \Big\{ \mathcal{J}(\mathbf{M}, \mathbf{dM}, \mathbf{A}) \text{ s.t. } (3) \Big\} \quad (4)$$

with

$$\begin{aligned} \mathcal{J}(\mathbf{M}, \mathbf{dM}, \mathbf{A}) = &\frac{1}{2}\left\|\mathbf{Y} - \mathbf{MA} - \boldsymbol{\Delta}\right\|_{\mathrm{F}}^2 + \alpha\Phi(\mathbf{A}) + \\ &\beta\Psi(\mathbf{M}) + \gamma\Upsilon(\mathbf{dM}) \end{aligned} \quad (5)$$

where the penalization parameters $\alpha, \beta, \gamma$ control the trade-off between the data fitting term $\frac{1}{2}\left\|\mathbf{Y} - \mathbf{MA} - \boldsymbol{\Delta}\right\|_{\mathrm{F}}^2$ and the

penalties $\Phi(\mathbf{A})$, $\Psi(\mathbf{M})$ and $\Upsilon(\mathbf{dM})$. In addition, we assume that the penalization functions are separable, leading to

$$\Phi(\mathbf{A}) = \sum_{n=1}^{N} \phi(\mathbf{a}_n) \tag{6}$$

$$\Psi(\mathbf{M}) = \sum_{\ell=1}^{L} \psi(\widetilde{\mathbf{m}}_\ell) \tag{7}$$

$$\Upsilon(\mathbf{dM}) = \sum_{n=1}^{N} \upsilon(\mathbf{dM}_n) \tag{8}$$

where $\widetilde{\mathbf{m}}_\ell$ denotes the $\ell$th row of $\mathbf{M}$ and $\phi$, $\psi$ and $\upsilon$ are non-negative differentiable convex functions. This assumption is used to decompose (4) into a collection of simpler sub-problems described in Section III. All these penalizations are described in the next paragraphs.

*1) Abundance penalization:* The abundance penalization $\Phi$ has been chosen to promote spatially smooth abundances as in [19]. More precisely, the abundance spatial smoothness penalization is expressed in matrix form as

$$\Phi(\mathbf{A}) = \frac{1}{2} \|\mathbf{AH}\|_{\mathrm{F}}^2 \tag{9}$$

where $\mathbf{H}$ is a matrix computing the differences between the abundances of a given pixel and those of its 4 nearest neighbors [19]. The resulting expression of $\phi$ is detailed in Appendix A.

*2) Endmember penalization:* As for $\Psi$, classical penalizations found in the literature consist of constraining the size of the simplex whose vertices are the endmember signatures. The volume criterion used in [20], [21] enables the volume exactly occupied by the $(K-1)$-simplex formed by the endmembers to be penalized. The mutual distance between the endmembers introduced in [22], [23] (which approximates the volume) has a similar purpose. Finally, if the endmembers are *a priori* close from available reference spectral signatures, a penalization on the distance between the endmembers and these signatures can be implemented. The expression of the distance between the endmembers and some reference spectral signatures, the mutual distance between the endmembers and the volume penalization are recalled in the following lines. For each penalization type, the corresponding expression of $\psi$ is given in Appendix A.

- The distance between the endmembers and some reference spectral signatures $\mathbf{M}_0$ is given by

$$\Psi(\mathbf{M}) = \frac{1}{2} \|\mathbf{M} - \mathbf{M}_0\|_{\mathrm{F}}^2 . \tag{10}$$

- The mutual distance between the endmembers is expressed in matrix form as

$$\Psi(\mathbf{M}) = \frac{1}{2} \sum_{i=1}^{K} \left( \sum_{\substack{j=1 \\ j \neq i}}^{K} \|\mathbf{m}_i - \mathbf{m}_j\|_2^2 \right). \tag{11}$$

- Under the pure pixel and linear mixture assumptions, the data points are enclosed in a $(K-1)$-simplex whose vertices are the endmembers [21]. Let $\mathbf{T}$ be the projection of $\mathbf{M}$ on the space spanned by the $K-1$ principal components of $\mathbf{Y}$. The expression of the volume of this subspace is

$$\mathcal{V}(\mathbf{T}) = \frac{1}{(K-1)!} \left| \det \begin{pmatrix} \mathbf{T} \\ \mathbf{1}_K^T \end{pmatrix} \right|.$$

To ensure the differentiability of the penalization with respect to $\mathbf{T}$, we propose to consider the following penalty

$$\Psi(\mathbf{M}) = \frac{1}{2} \mathcal{V}^2(\mathbf{T}). \tag{12}$$

*3) Variability penalization:* The variability penalizing function $\Upsilon$ has been designed to limit the norm of the spectral variability. Indeed, it is legitimate to penalize the energy of the perturbation matrices $\mathbf{dM}_n$ in order to obtain a reasonable endmember variability. In this paper, we propose to consider the following penalty (having the advantage to be differentiable with respect to $\mathbf{dM}_n$)

$$\Upsilon(\mathbf{dM}) = \frac{1}{2} \|\mathbf{dM}\|_{\mathrm{F}}^2 = \frac{1}{2} \sum_{n=1}^{N} \|\mathbf{dM}_n\|_{\mathrm{F}}^2 . \tag{13}$$

To the best of our knowledge, no specific information regarding the spatial distribution of the variability is available in the remote sensing literature so far. We have consequently preferred not to include any additional regularization on $\mathbf{dM}$. However, any spatial penalization satisfying the assumptions given in Paragraph II-B can be added when necessary (e.g., a group-Lasso $\ell_{2,1}$ penalization to promote spatial sparsity of the variability term $\mathbf{dM}$).

**Algorithm 1:** PLMM-unmixing: global algorithm.

**Data**: $\mathbf{Y}, \mathbf{A}^{(0)}, \mathbf{M}^{(0)}, \mathbf{dM}^{(0)}$
**Result**: $\mathbf{A}, \mathbf{M}, \mathbf{dM}$
**begin**
  $k \leftarrow 1$;
  **while** *stopping criterion not satisfied* **do**
    a $\mathbf{A}^{(k)} \leftarrow \arg\min_{\mathbf{A}} \mathcal{J}\left(\mathbf{M}^{(k-1)}, \mathbf{dM}^{(k-1)}, \mathbf{A}\right)$ ;
    b $\mathbf{M}^{(k)} \leftarrow \arg\min_{\mathbf{M}} \mathcal{J}\left(\mathbf{M}, \mathbf{dM}^{(k-1)}, \mathbf{A}^{(k)}\right)$ ;
    c $\mathbf{dM}^{(k)} \leftarrow \arg\min_{\mathbf{M}} \mathcal{J}\left(\mathbf{M}^{(k)}, \mathbf{dM}, \mathbf{A}^{(k)}\right)$ ;
    $k \leftarrow k+1$;
  $\mathbf{A} \leftarrow \mathbf{A}^{(k)}$;
  $\mathbf{M} \leftarrow \mathbf{M}^{(k)}$;
  $\mathbf{dM} \leftarrow \mathbf{dM}^{(k)}$;

## III. An ADMM-based algorithm

Since the problem (4) is not convex, a minimization strategy similar to [12] has been adopted. Precisely, the cost function $\mathcal{J}$ is successively minimized with respect to each variable $\mathbf{A}, \mathbf{M}$ and $\mathbf{dM}$ until a stopping criterion is satisfied. The assumptions made on the penalization functions $\Phi, \Psi, \Upsilon$ in Section II allow the global optimization problem to be divided into a collection of strictly convex sub-problems. These sub-problems have

the nice property to involve differentiable functions simplifying their resolution. Having introduced appropriate splitting variables to account for the constraints, these sub-problems are finally solved using ADMM steps admitting closed-form expressions due to the separability assumption. The three minimization steps considered in this algorithm present a highly similar structure. The details are reported in Appendix B to facilitate the reading of this paper.

### A. ADMM: general principle

The ADMM is a technique combining the benefits of augmented Lagrangian and dual decomposition methods to solve constrained optimization problems [24]. More precisely, the method consists of solving the original optimization problem by successively minimizing the cost function of interest with respect to each variable. The following elements (extracted from [24]) recall a general formulation of the problem. Given $f : \mathbb{R}^p \to \mathbb{R}^+$, $g \in \mathbb{R}^m \to \mathbb{R}^+$, $\mathbf{A} \in \mathbb{R}^{n\times p}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$, consider the general optimization problem

$$\min_{\mathbf{x},\mathbf{z}} \Big\{ f(\mathbf{x}) + g(\mathbf{z}) \Big| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \Big\}. \tag{14}$$

The scaled augmented Lagrangian associated with this problem can be written

$$\mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{u}) = f(\mathbf{x}) + g(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \mathbf{u}\right\|_2^2$$

where $\rho > 0$. Denote as $\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)}$ and $\mathbf{u}^{(k+1)}$ the primal variables and the dual variable at iteration $k+1$ of the algorithm

$$\begin{aligned}
\mathbf{x}^{(k+1)} &\in \arg\min_{\mathbf{x}} \mathcal{L}_\rho\left(\mathbf{x}, \mathbf{z}^{(k)}, \mathbf{u}^{(k)}\right) \\
\mathbf{z}^{(k+1)} &\in \arg\min_{\mathbf{z}} \mathcal{L}_\rho\left(\mathbf{x}^{(k+1)}, \mathbf{z}, \mathbf{u}^{(k)}\right) \\
\mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} + \mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{z}^{(k+1)} - \mathbf{c}.
\end{aligned}$$

The ADMM consists in successively minimizing $\mathcal{L}_\rho$ with respect to $\mathbf{x}, \mathbf{z}$ and $\mathbf{u}$. A classical stopping criterion involves the primal and dual residuals at iteration $k+1$ (see [24, p. 19]): the procedure is iterated until

$$\left\|\mathbf{r}^{(k)}\right\|_2 \leq \varepsilon^{\text{pri}} \quad \text{and} \quad \left\|\mathbf{s}^{(k)}\right\|_2 \leq \varepsilon^{\text{dual}} \tag{15}$$

where the primal and dual residuals at iteration $k+1$ are respectively given by

$$\mathbf{r}^{(k+1)} = \mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{z}^{(k+1)} - \mathbf{c} \tag{16}$$

$$\mathbf{s}^{(k+1)} = \rho\mathbf{A}^T\mathbf{B}\left(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\right) \tag{17}$$

and

$$\varepsilon^{\text{pri}} = \sqrt{p}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{ \left\|\mathbf{A}\mathbf{x}^{(k)}\right\|_2^2, \left\|\mathbf{B}\mathbf{z}^{(k)}\right\|_2^2, \|\mathbf{c}\|_2^2 \right\} \tag{18}$$

$$\varepsilon^{\text{dual}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \left\|\mathbf{A}^T\mathbf{y}^{(k)}\right\|_2^2. \tag{19}$$

**Algorithm 2:** ADMM optimization w.r.t. $\mathbf{A}$ (step (a)).

**Data**: $\mathbf{Y}, \mathbf{A}^{(0)}, \mathbf{M}^{(0)}, \varepsilon_{\text{pri}}, \varepsilon_{\text{dual}}, \tau^{\text{incr}}, \tau^{\text{decr}}, \mu_n^{(\mathbf{A})(0)}$
**Result**: $\mathbf{A}$
**for** $n = 1$ **to** $N$ **do**
  $k \leftarrow 1$;
  $\boldsymbol{\lambda}_n^{(\mathbf{A})(0)} = \mathbf{0}$;
  $\mathbf{w}_n^{(\mathbf{A})(0)} = \mathbf{0}$;
  **while** *stopping criterion not satisfied* **do**
    $\mathbf{a}_n^{(k)} \leftarrow$
    $\arg\min_{\mathbf{a}_n} \mathcal{L}_{\mu_n^{(\mathbf{A})(k-1)}}\left(\mathbf{a}_n, \mathbf{w}_n^{(\mathbf{A})(k-1)}, \boldsymbol{\lambda}_n^{(\mathbf{A})(k-1)}\right)$;
    $\mathbf{w}_n^{(\mathbf{A})(k)} \leftarrow$
    $\arg\min_{\mathbf{w}_n^{(\mathbf{A})}} \mathcal{L}_{\mu_n^{(\mathbf{A})(k-1)}}\left(\mathbf{a}_n^{(k)}, \mathbf{w}_n^{(\mathbf{A})}, \boldsymbol{\lambda}_n^{(\mathbf{A})(k-1)}\right)$;
    $\boldsymbol{\lambda}_n^{(\mathbf{A})(k)} \leftarrow \boldsymbol{\lambda}_n^{(\mathbf{A})(k-1)} + \mathbf{Q}\mathbf{a}_n^{(k)} + \mathbf{w}_n^{(\mathbf{A})(k)} - \mathbf{s}$ ;
    $\mu_n^{(\mathbf{A})(k)} \leftarrow \text{Update}\left(\mu_n^{(\mathbf{A})(k-1)}\right)$ using (20) ;
    $k \leftarrow k+1$ ;
  $\mathbf{a}_n \leftarrow \mathbf{a}_n^{(k)}$;

Finally, the parameter $\rho$ can be adjusted using the rule described in [24, p. 20]

$$\rho^{(k+1)} = \begin{cases} \tau^{\text{incr}}\rho^{(k)} & \text{if } \left\|\mathbf{r}^{(k)}\right\|_2 > \mu\left\|\mathbf{s}^{(k)}\right\|_2 \\ \rho^{(k)}/\tau^{\text{decr}} & \text{if } \left\|\mathbf{s}^{(k)}\right\|_2 > \mu\left\|\mathbf{r}^{(k)}\right\|_2 \\ \rho^{(k)} & \text{otherwise.} \end{cases} \tag{20}$$

Note that this parameter adjustment does not alter the ADMM convergence as long as it is performed finitely many times.

### B. Optimization with respect to A

With the assumptions made in paragraph II-B, optimizing the cost function $\mathcal{J}$ with respect to $\mathbf{A}$ under the constraints (3) is equivalent to solving the following problems

$$\mathbf{a}_n^* = \arg\min_{\mathbf{a}_n} \left\{ \begin{array}{c} \frac{1}{2}\left\|\mathbf{y}_n - (\mathbf{M} + \mathbf{d}\mathbf{M}_n)\mathbf{a}_n\right\|_2^2 + \alpha\phi(\mathbf{a}_n) \\ \text{s.t.} \quad \mathbf{a}_n \succeq \mathbf{0}_K, \quad \mathbf{a}_n^T\mathbf{1}_K = 1 \end{array} \right\}. \tag{21}$$

After introducing the splitting variables $\mathbf{w}_n^{(\mathbf{A})} \in \mathbb{R}^K$ for $n = 1, \ldots, N$ such that

$$\underbrace{\begin{pmatrix} \mathbf{I}_K \\ \mathbf{1}_K^T \end{pmatrix}}_{\mathbf{Q}} \mathbf{a}_n + \underbrace{\begin{pmatrix} -\mathbf{I}_K \\ \mathbf{0}_K^T \end{pmatrix}}_{\mathbf{R}} \mathbf{w}_n = \underbrace{\begin{pmatrix} \mathbf{0}_K \\ 1 \end{pmatrix}}_{\mathbf{s}} \tag{22}$$

the resulting scaled augmented Lagrangian is expressed as

$$\begin{aligned}
\mathcal{L}_{\mu_n^{(\mathbf{A})}}\left(\mathbf{a}_n, \mathbf{w}_n^{(\mathbf{A})}, \boldsymbol{\lambda}_n^{(\mathbf{A})}\right) &= \frac{1}{2}\left\|\mathbf{y}_n - (\mathbf{M} + \mathbf{d}\mathbf{M}_n)\mathbf{a}_n\right\|_2^2 \\
&+ \frac{\mu_n^{(\mathbf{A})}}{2}\left\|\mathbf{Q}\mathbf{a}_n + \mathbf{R}\mathbf{w}_n^{(\mathbf{A})} - \mathbf{s} + \boldsymbol{\lambda}_n^{(\mathbf{A})}\right\|_2^2 \\
&+ \alpha\phi(\mathbf{a}_n) + \mathcal{I}_{\mathcal{S}_{K,1}^+}\left(\mathbf{w}_n^{(\mathbf{A})}\right)
\end{aligned} \tag{23}$$

with $\mu_n^{(\mathbf{A})} > 0$. The resulting algorithm (step a of Algo. 1) is detailed in Algo. 2, and the solution to each sub-problem is given in Appendix B.

### C. Optimization with respect to $\boldsymbol{M}$

Similarly to Paragraph III-B, optimizing $\mathcal{J}$ with respect to $\mathbf{M}$ under the constraint (3) is equivalent to solving

$$\widetilde{\mathbf{m}}_\ell^* = \arg\min_{\widetilde{\mathbf{m}}_\ell} \left\{ \begin{array}{c} \frac{1}{2}\left\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\right\|_2^2 + \beta\psi(\widetilde{\mathbf{m}}_\ell) \\ \text{s.t.} \quad \text{for} \quad n = 1,\ldots,N \\ \widetilde{\mathbf{m}}_\ell \succeq \mathbf{0}_K^T, \quad \widetilde{\mathbf{m}}_\ell + \widetilde{\mathbf{dm}}_{n,\ell} \succeq \mathbf{0}_K^T \end{array} \right\} \tag{24}$$

where $\widetilde{\mathbf{m}}_\ell$ denotes the $\ell$th row of $\mathbf{M}$. Introducing the splitting variables $\mathbf{W}_\ell^{(\mathbf{M})} \in \mathbb{R}^{(N+1)\times K}$ for $\ell = 1,\ldots,L$ such that

$$\underbrace{\begin{pmatrix} 1 \\ \mathbf{1}_N \end{pmatrix}}_{\mathbf{e}} \widetilde{\mathbf{m}}_\ell - \mathbf{W}_\ell^{(\mathbf{M})} = -\underbrace{\begin{bmatrix} \mathbf{0}_K^T \\ \begin{pmatrix} \widetilde{\mathbf{dm}}_{1,\ell} \\ \vdots \\ \widetilde{\mathbf{dm}}_{N,\ell} \end{pmatrix} \end{bmatrix}}_{\mathbf{F}_\ell} \tag{25}$$

the associated scaled augmented Lagrangian can be written

$$\begin{aligned} \mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \mathbf{W}_\ell^{(\mathbf{M})}, \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right) &= \frac{1}{2}\left\|\widetilde{\mathbf{y}}_l - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\right\|_2^2 \\ &+ \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\mathbf{e}\widetilde{\mathbf{m}}_\ell - \mathbf{W}_\ell^{(\mathbf{M})} + \mathbf{F}_\ell + \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right\|_{\mathrm{F}}^2 \\ &+ \beta\psi(\widetilde{\mathbf{m}}_\ell) + \mathcal{I}_{\mathcal{S}_{N+1,K}^+}\left(\mathbf{W}_\ell^{(\mathbf{M})}\right) \end{aligned} \tag{26}$$

with $\mu_\ell^{(\mathbf{M})} > 0$. The resulting algorithm (step b of Algo. 1) is similar to Algo. 2. The solution to the optimization problems depends on the selected endmember penalizing function $\Psi$ chosen in paragraph II-B2 (see Appendix B for more details).

### D. Optimization with respect to $\boldsymbol{dM}$

Finally, optimizing $\mathcal{J}$ with respect to $\mathbf{dM}$ under the constraint (3) is equivalent to solving the sub-problems

$$\mathbf{dM}_n^* = \arg\min_{\mathbf{dM}_n} \left\{ \begin{array}{c} \frac{1}{2}\left\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\right\|_2^2 \\ +\gamma\upsilon(\mathbf{dM}_n) \\ \text{s.t.} \quad \mathbf{M} + \mathbf{dM}_n \succeq \mathbf{0}_{L,K} \end{array} \right\}. \tag{27}$$

Introducing the splitting variables $\mathbf{W}_n^{(\mathbf{dM})} = \mathbf{M} + \mathbf{dM}_n$ for $n = 1,\ldots,N$, the resulting scaled augmented Lagrangian is given by

$$\begin{aligned} \mathcal{L}_{\mu_n^{(\mathbf{dM})}}\left(\mathbf{dM}_n, \mathbf{W}_n^{(\mathbf{dM})}, \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right) &= \frac{1}{2}\left\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_\mathbf{n}\right\|_2^2 \\ &+ \frac{\mu_n^{(\mathbf{dM})}}{2}\left\|\mathbf{dM}_n + \mathbf{M} - \mathbf{W}_n^{(\mathbf{dM})} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right\|_{\mathrm{F}}^2 \\ &+ \gamma\upsilon(\mathbf{dM}_n) + \mathcal{I}_{\mathcal{S}_{L,K}^+}\left(\mathbf{W}_n^{(\mathbf{dM})}\right) \end{aligned} \tag{28}$$

with $\mu_n^{(\mathbf{dM})} > 0$. The resulting algorithm (step c of Algo. 1) is similar to Algo. 2. The solution to these problems is given in Appendix B.

The optimization procedures detailed above are performed sequentially until the stopping criterion is satisfied. The next sections evaluate the performance of the resulting unmixing strategy via several experiments conducted on synthetic and real data.

### E. Convergence and computational cost

The alternating scheme proposed in Alg. 1 is nothing but a block coordinate descent descent (BCD) which is guaranteed to converge to a stationary point of the objective function $\mathcal{J}$ as long as each sub-problem is exactly minimized [25, Proposition 2.7.1]. Besides, the sub-problems tackled in III-B, III-C and III-D are strongly convex, hence the convergence of the ADMM steps toward the unique minimum of each independent sub-problem when the augmented Lagrangian parameter has a constant value (see for instance [24]). The same convergence result applies to the ADMM with the parameter adjustment introduced in Paragraph III-A as long as the parameter is updated finitely many times [24]. We may however mention that the proximal alternating linearized minimization (PALM) [26] could also be directly applied to the considered problem with a rigorous convergence proof based on the Kurdyka-Łojasiewicz property. This alternative work has been presented in [27].

Considering the significant number of unknown parameters and the simple expression of the ADMM updates detailed in Appendix B, we can note that the computational cost is dominated by matrix products, yielding an overall $\mathcal{O}(LK^2N)$ computational cost.

## IV. Experiment with synthetic data

This section considers four images of size $128 \times 64$ acquired in $413$ bands. Each image corresponds to a mixture of $K$ endmembers with $K \in \{3, 6\}$ in presence or absence of pure pixels (the absence of pure pixels is considered to evaluate the algorithm performance in a very challenging scenario). The synthetic linear mixtures have been corrupted by additive white Gaussian noise to ensure the signal-to-noise ratio is SNR $= 30$dB. Since no accepted variability model is available in the literature, we propose the following generation procedure to introduce controlled spectral variability. The corrupted endmembers involved in the mixture (see Fig. 2) have been generated using the product of reference endmembers with randomly drawn piece-wise affine functions, providing realistic perturbed endmembers as represented in Fig. 1. For a given variability coefficient $c_{\text{var}} > 0$ fixed by the user, the parameters $\xi_i$, $i \in \{1, 2, 3\}$ and $L_{\text{break}} \in \{1,\ldots,L\}$ introduced in Fig. 1 have been generated as follows

$$\xi_i \sim \mathcal{U}_{[1-c_{\text{var}}/2, 1+c_{\text{var}}/2]},\ i \in \{1, 2, 3\} \tag{29}$$

$$L_{\text{break}} = \lfloor L/2 + \lfloor LU/3 \rfloor \rfloor,\ U \sim \mathcal{N}(0, 1) \tag{30}$$

where $\lfloor\cdot\rfloor$ denotes the floor operator. The synthetic data used in the proposed experiments have been generated with a value of $c_{\text{var}}$ that is lower in the upper half of the image ($c_{\text{var}} = 0.1$) than in the lower half ($c_{\text{var}} = 0.25$). Some instances of the corresponding perturbed endmember spectra are depicted in Fig. 2. Note that different affine functions have been considered for different endmembers and different pixels.

### A. State-of-the-art methods

The results of the proposed algorithm have been compared to those obtained with two classical linear unmixing methods

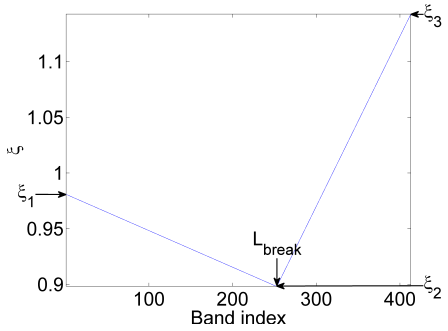



Fig. 1. Example of a randomly-generated affine function used to generate the synthetically perturbed endmembers.

(VCA/FCLS, SISAL/FCLS) and two variability accounting algorithms (AEB, FDNS). These methods are recalled below with their most relevant implementation details.

1) Classical unmixing methods (no variability)
   - VCA/FCLS: the endmembers are first extracted using the vertex component analysis [28]. The abundances are then estimated for each pixel using the fully constrained least squares (FCLS) algorithm [29]);
   - SISAL/FCLS: the endmembers are first extracted using the simplex identification via split augmented Lagrangian [9]. The tolerance for the stopping rule has been set to $10^{-2}$ and VCA has been used as an initialization step. The abundances are then estimated for each pixel using FCLS.
2) Variability accounting unmixing methods
   - AEB [5], [30], [31]: the size of the bundles is equal to 25% of the total pixel number. The endmembers and abundance are estimated using VCA/FCLS;
   - FDNS [6]: the endmembers and abundances are estimated by VCA/FCLS ;
   - Proposed method (BCD/ADMM): endmembers and abundances have been initialized with VCA/FCLS estimates. Note that VCA/FCLS is a method assuming that there are pure pixels in the image, which can be problematic in case the imaged scene does not satisfy this assumption. The variability matrices have been initialized with all their entries equal to $eps$[1]. The algorithm is stopped when the relative difference between two successive values of the objective function is less than $10^{-3}$.

Different penalization combinations have been compared for the proposed method. The abbreviations *ss*, *mv* and *vca* are used for *spatially smooth*, *minimum volume* and *minimum distance to VCA* in the following. The absence of any additional abbreviation means that the method does not include any abundance or endmember penalization term.

The performance of the algorithm has been assessed in terms of endmember estimation using the average spectral angle mapper (aSAM)

$$\text{aSAM}(\mathbf{M}) = \frac{1}{K}\sum_{k=1}^{K} \arccos\left(\frac{\langle \mathbf{m}_k | \widehat{\mathbf{m}}_k \rangle}{\|\mathbf{m}_k\|_2 \|\widehat{\mathbf{m}}_k\|_2}\right)$$

as well as in terms of abundance and perturbation estimation by global mean square errors (GMSEs)

$$\text{GMSE}(\mathbf{A}) = \frac{1}{KN}\|\mathbf{A} - \widehat{\mathbf{A}}\|_{\text{F}}^2$$

$$\text{GMSE}(\mathbf{dM}) = \frac{1}{NLK}\sum_{n=1}^{N}\|\mathbf{dM}_n - \widehat{\mathbf{dM}}_n\|_{\text{F}}^2.$$

As a measure of fit, the following reconstruction error (RE)

[1]MATLAB constant $eps = 2.22 \times 10^{-16}$.

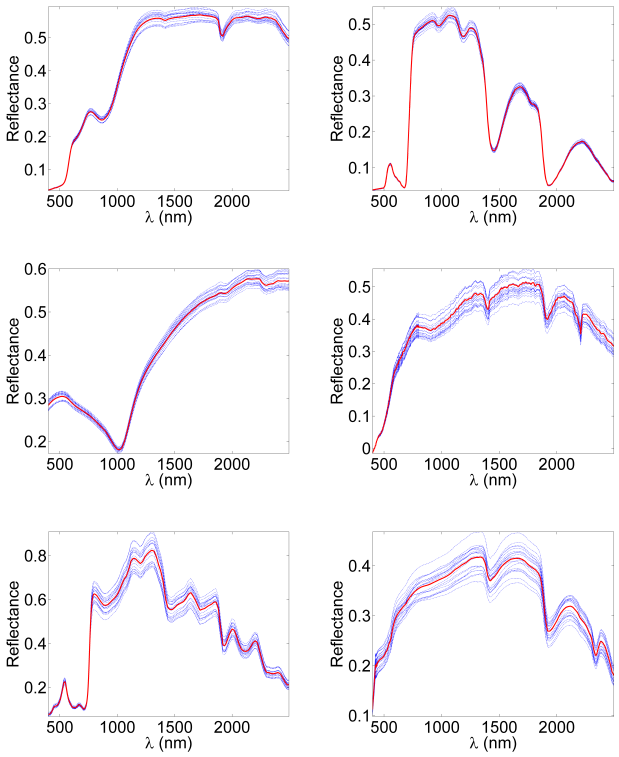
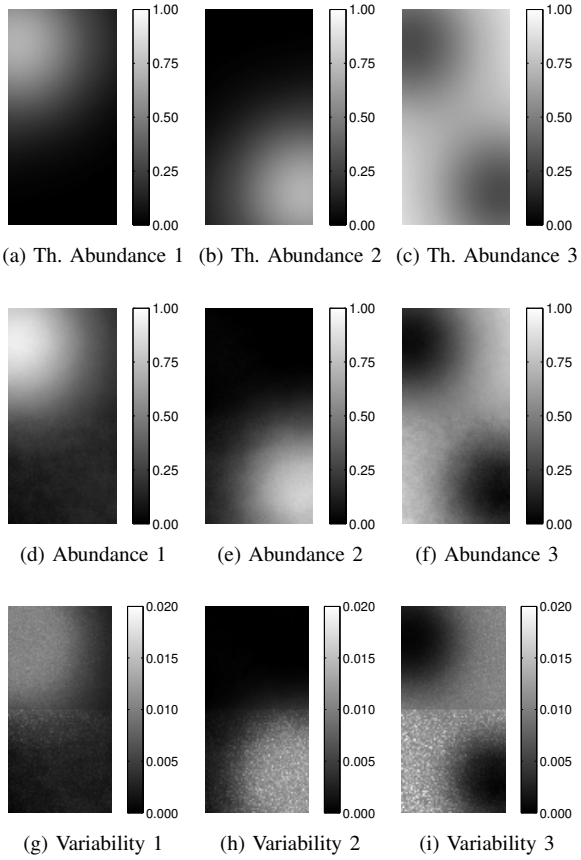



Fig. 2. Reference endmembers (red lines) and 20 corresponding instances under spectral variability (blue dotted lines) involved in the synthetic data experiments.

TABLE II
ADMM PARAMETERS.

| | Synthetic data | Real data |
|---|---|---|
| $\tau^{\text{incr}}$ | 1.1 | 1.1 |
| $\tau^{\text{decr}}$ | 1.1 | 1.1 |
| $\mu$ | 10 | 10 |
| $\mu_n^{(\mathbf{A})(0)}$ | $10^{-4}$ | $10^{-4}$ |
| $\mu_\ell^{(\mathbf{M})(0)}$ | $10^{-8}$ | $10^{-8}$ |
| $\mu_n^{(\mathbf{dM})(0)}$ | $10^{-4}$ | $10^{-4}$ |
| $\varepsilon^{\text{abs}}$ | $10^{-1}$ | $10^{-2}$ |
| $\varepsilon^{\text{rel}}$ | $10^{-4}$ | $10^{-4}$ |




(a) Th. Abundance 1 (b) Th. Abundance 2 (c) Th. Abundance 3

(d) Abundance 1 (e) Abundance 2 (f) Abundance 3

(g) Variability 1 (h) Variability 2 (i) Variability 3

Fig. 3. True abundances (fig. 3a to 3c) and the ssmvBCD/ADMM-estimations (fig. 3d to 4c) – obtained with a synthetic dataset (no pure pixels, $K = 3$). The spatial distribution of the variability with respect to each endmember is presented in terms of energy ($\frac{1}{\sqrt{L}}\|\mathbf{dm}_{n,k}\|_2$ for the $k$th endmember in the $n$th pixel) for visualization purpose in Figs. 3g to 3i.

has been also considered

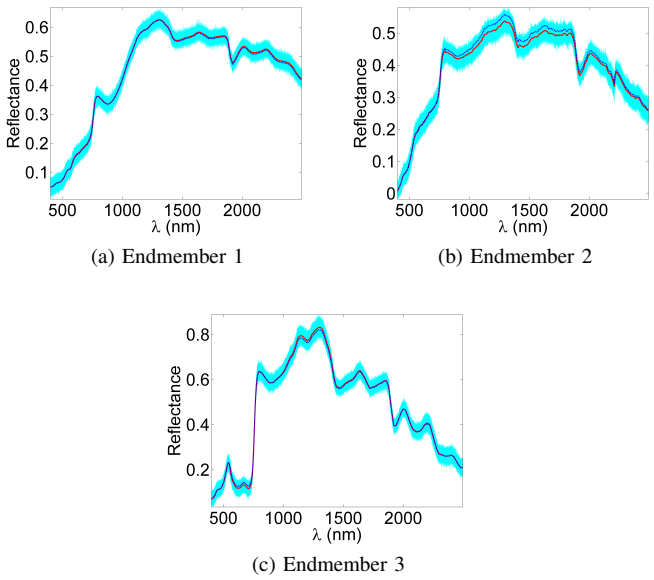
$$\mathrm{RE} = \frac{1}{LN}\left\|\mathbf{Y} - \widehat{\mathbf{Y}}\right\|_{\mathrm{F}}^2$$

where $\widehat{\mathbf{Y}}$ is the matrix formed of the pixels reconstructed using the parameters estimated by the algorithm.

### B. Results

The parameters used for the ADMM algorithms are detailed in Table II, and the values chosen by cross-validation for $\alpha$, $\beta$ and $\gamma$ are reported in Table III and IV. The performance measures returned by the unmixing methods are provided in Table III for the datasets containing pure pixels, and in Table IV for images without pure pixels, leading to the following conclusions.

- The proposed method is robust to the absence of pure pixels;
- The proposed method provides competitive results in terms of aSAM while allowing endmember variability to be estimated for each endmember in each pixel;
- For most simulation scenarios, the abundance MSEs and the REs are lower than the MSEs and REs resulting from state-of-the-art methods;




(a) Endmember 1 (b) Endmember 2

(c) Endmember 3

Fig. 4. Endmember estimations obtained on synthetic data in absence of pure pixels (cf. Figs. 3 for the abundance estimations). The ssmvBCD/ADMM-estimated endmembers (red lines) are given with typical examples of the estimated variability (cyan dotted lines). The VCA endmembers are given in blue dotted lines for comparison.

- The proposed method is computationally more expensive than existing algorithms.

We can note that the smoothness penalization on the abundances proves to be particularly appropriate in this experiment. Moreover, an increasing number of endmembers implies a loss of estimation performance. This result can be expected since VCA/FCLS algorithm is used as an initialization step.

Finally, the variability captured by the proposed model is presented in Figs. 3 and 4 for three endmembers: the difference between the variability intensities detected in the upper and the lower part of the scene is due to the different variability coefficients applied to these areas, thus illustrating the consistency of the proposed method.

## V. Experiment with real data

### A. Description of the datasets

The proposed algorithm has been applied to real hyperspectral datasets obtained by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). The first scene was acquired over Moffett Field, CA, in 1997. Water absorption bands were removed from the 224 spectral bands, leaving 189 exploitable spectral bands. The scene of interest ($50 \times 50$) is partly composed of a lake and a coastal area.

The second scene is a $190 \times 250$ image extracted from the well-known Cuprite dataset[2]. The number of spectral bands is 189 after removing the water-absorption and low SNR bands. Many works previously conducted on this image provide reference abundance estimation maps.

[2]The Moffett and Cuprite images are available at http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, and http://aviris.jpl.nasa.gov/

TABLE III
SIMULATION RESULTS FOR SYNTHETIC DATA IN PRESENCE OF PURE PIXELS (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$, $\gamma = 1$).

| | $K = 3$, $(\alpha, \beta) = (1.4, 2.5 \times 10^{-5})$ | | | | | $K = 6$, $(\alpha, \beta) = (0.37, 5.1 \times 10^{-4})$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | aSAM($\mathbf{M}$) (°) | GMSE($\mathbf{A}$) | GMSE($\mathbf{dM}$) | RE | time (s) | aSAM($\mathbf{M}$) (°) | GMSE($\mathbf{A}$) | GMSE($\mathbf{dM}$) | RE | time (s) |
| VCA/FCLS | 6.0038 | 3.80 | / | 7.56 | **1** | 6.3313 | 2.24 | / | 2.92 | **1** |
| SISAL | **5.2665** | 3.08 | / | 3.35 | 2 | **3.8365** | 3.05 | / | 2.25 | 3 |
| FDNS | 6.0038 | 3.79 | / | 7.56 | 4 | 6.3313 | 2.22 | / | 2.92 | 5 |
| AEB | 5.6971 | **2.07** | / | 3.50 | 52 | 5.7017 | **1.31** | / | 2.40 | 142 |
| BCD/ADMM | 5.9910 | 3.51 | **4.00** | **0.20** | 92 | 6.2965 | 1.59 | **2.93** | **0.05** | 230 |
| ssBCD/ADMM | 5.7765 | 3.15 | 4.25 | 0.23 | 422 | 6.0304 | 1.44 | 2.97 | 0.07 | 848 |
| ssmvBCD/ADMM | 5.4390 | 3.01 | 4.25 | 0.25 | 530 | 6.3397 | 1.42 | 2.97 | 0.07 | 603 |

TABLE IV
SIMULATION RESULTS FOR SYNTHETIC DATA IN ABSENCE OF PURE PIXELS (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$, $\gamma = 1$).

| | $K = 3$, $(\alpha, \beta) = (24.5, 4.2 \times 10^{-9})$ | | | | | $K = 6$, $(\alpha, \beta) = (0.71, 4.8 \times 10^{-4})$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | aSAM($\mathbf{M}$) (°) | GMSE($\mathbf{A}$) | GMSE($\mathbf{dM}$) | RE | time (s) | aSAM($\mathbf{M}$) (°) | GMSE($\mathbf{A}$) | GMSE($\mathbf{dM}$) | RE | time (s) |
| VCA/FCLS | 5.0639 | 2.07 | / | 2.66 | **1** | 6.5530 | 2.52 | / | 2.82 | **4** |
| SISAL | 4.4318 | 2.16 | / | 2.56 | 2 | 6.0431 | **1.63** | / | 2.02 | 5 |
| FDNS | 5.0639 | 2.06 | / | 2.66 | 3 | 6.5530 | 2.53 | / | 2.82 | 7 |
| AEB | 5.1104 | 2.11 | / | 2.66 | 33 | **6.0016** | 1.78 | / | 1.85 | 208 |
| BCD/ADMM | 5.2480 | 2.13 | **3.81** | **0.25** | 140 | 6.2785 | 2.14 | 3.33 | 0.30 | 3041 |
| ssBCD/ADMM | **4.1549** | **1.44** | 4.36 | 0.38 | 1263 | 6.2763 | 1.74 | **3.04** | **0.076** | 1527 |
| ssmvBCD/ADMM | 5.0584 | 1.94 | 4.59 | 0.47 | 1667 | 6.3207 | 1.67 | 3.05 | 0.08 | 795 |

The parameters used for the proposed approach are identical to those used for the experiments with synthetic data (see Table II). The only difference is that the algorithm has been stopped when the relative difference between two successive values of the objective function is less than $10^{-2}$. This value has been chosen to obtain a compromise between the estimation accuracy and the computational cost implied. The values selected by cross-validation for $\alpha$, $\beta$ and $\gamma$ are given in Table V.

TABLE V
EXPERIMENT RESULTS CONDUCTED ON REAL DATA [SSBCD/ADMM FOR MOFFETT WITH $(\alpha, \beta) = (0.05, 0)$, SSVCABCD/ADMM FOR CUPRITE WITH $(\alpha, \beta) = (0.014, 404)$, RE $\times 10^{-4}$,$\gamma = 1$].

| | Moffett | | Cuprite | |
|---|---|---|---|---|
| | RE | time (s) | RE | time (s) |
| VCA/FCLS | 2.50 | **0.4** | 3.69 | **9.9** |
| SISAL | 1.12 | 30 | 2.16 | 15 |
| FDNS | 2.69 | 1 | 3.69 | 11 |
| AEB | 6.25 | 10 | 0.40 | 615 |
| BCD/ADMM | **0.18** | 144 | **0.23** | 1.9e4 |

### B. Results

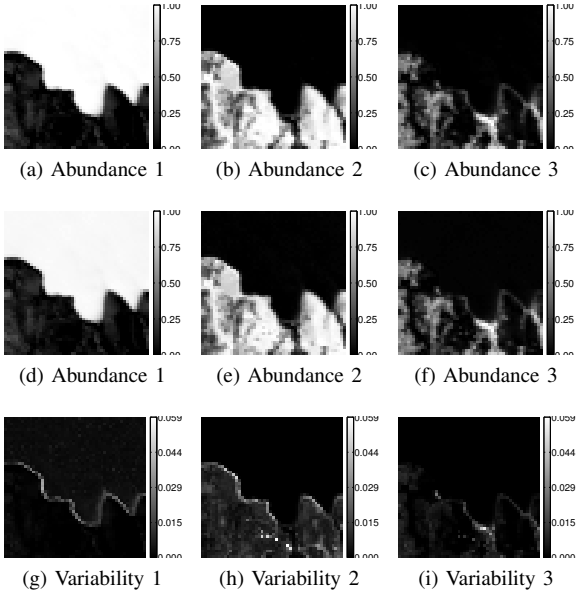
The unmixing performance are reported in Table V. For the Moffett image, the variability detected by the proposed algorithm is displayed in Figs. 5 and 6. The variability seems to be more significant on the coastal area where the mixture is not appropriately described by a linear model. The potential non-linearities usually observed close to the coastal areas [32]–[34] are interpreted as variability in the proposed method, which tends to corroborate its consistency. Note that the advantage of the proposed method is that it does not require to consider a sophisticated non-linear model accounting for interactions between the different endmembers as in [32], [34], [35]. Conversely, all deviations from the LMM are contained in the variability components $\mathbf{dM}_{n,k}$. We can also note that the variability peaks observed in Fig. 6 are a clear indication that several corrupted spectral bands have not been removed prior to the unmixing process.

The results obtained for the Cuprite scene are reported in Figs. 7, 8 and 9. Comparing our results with those of [28], we visually found out that some similar endmembers that were identified as different signatures by VCA for $K = 14$ [28] are interpreted as multiple instances of single endmembers in our setting ($K = 10$). The identification is given in Figs. 7 and 9. Fig. 8 shows that the algorithm captured a significant variability level in the pixels where many different endmembers are detected, which reveals that the spectral mixture may not be strictly linear in these pixels.

## VI. CONCLUSION AND FUTURE WORK

This paper introduced a new linear mixing model accounting for spatial and spectral endmember variabilities. The proposed model extended the classical LMM by including an additive spatially varying perturbation matrix that can capture endmember variabilities. The resulting unmixing problem was solved by alternating marginal minimizations of an appropriately regularized cost function, each minimization being performed by an ADMM algorithm. Simulations conducted on synthetic and real data enabled the interest of the proposed solution to be appreciated. Indeed, the proposed method compared favorably with state-of-the-art approaches while providing a relevant variability estimation. The choice of the penalization

parameters $\alpha, \beta$ and $\gamma$ was performed by cross validation. We think that it would be interesting to develop automatic strategies for estimating these parameters. Finally, due to the significant number of unknown parameters, the proposed method is not intended to be applied to very large images. The proposed approach can be applied as a complementary tool when analyzing small hyperspectral images *a priori* believed to be affected by a non-negligible variability level. Decreasing the computational complexity of the algorithm introduced in this work is clearly an interesting prospect.

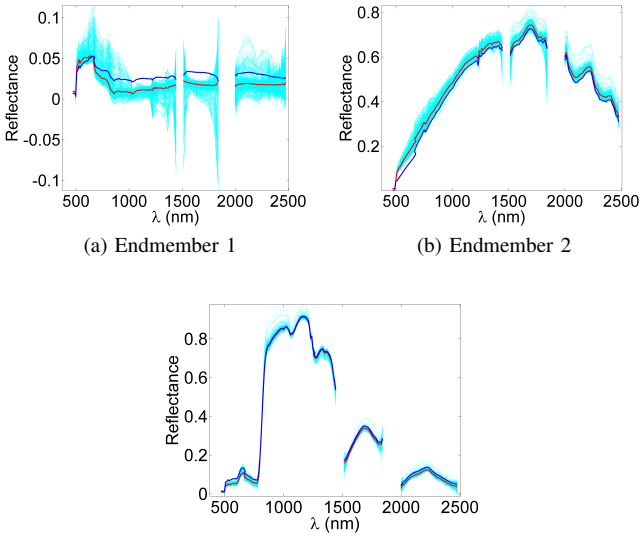


(a) Abundance 1 (b) Abundance 2 (c) Abundance 3

(d) Abundance 1 (e) Abundance 2 (f) Abundance 3

(g) Variability 1 (h) Variability 2 (i) Variability 3

Fig. 5. ssmvBCD/ADMM abundance estimations (Moffett scene). VCA/FCLS results are shown in Figs. 5a to 5c whereas Figs. 5d to 6c are for the ssmvBCD/ADMM. The spatial distribution of the variability with respect to each endmember is presented in terms of energy ($\frac{1}{\sqrt{L}}\|\mathbf{dm}_{n,k}\|_2$ for the $k$th endmember in the $n$th pixel) for visualization purpose.




(a) Endmember 1 (b) Endmember 2

(c) Endmember 3

Fig. 6. ssmvBCD/ADMM endmembers (Moffett scene). The ssmvBCD/ADMM-estimated endmembers (red lines) are plotted with the VCA endmembers (blue lines) for comparison, and typical examples of the estimated variability are given in cyan dotted lines.

## Appendix A
## Constraints and Penalization Terms

### A. Abundance penalization: spatial smoothness

The abundance smoothness is expressed in matrix form as

$$\Phi(\mathbf{A}) = \frac{1}{2}\|\mathbf{AH}\|_{\mathrm{F}}^2 \tag{31}$$

where $\mathbf{H}$ denotes the matrix computing the differences between the abundances of a given pixel and the respective abundances of its 4 neighbors

$$\mathbf{H} = \left[\begin{array}{c|c|c|c} \mathbf{H}_{\leftarrow} & \mathbf{H}_{\rightarrow} & \mathbf{H}_{\uparrow} & \mathbf{H}_{\downarrow} \end{array}\right] \in \mathbb{R}^{N\times 4N}.$$

For $h = 1, \ldots, H$, we introduce

$$\mathbf{H}_h = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 & -1 \\ 0 & \ldots & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{W\times W}$$

$$\tilde{\mathbf{H}}_h = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & -1 & 0 \end{pmatrix} \in \mathbb{R}^{W\times W}.$$

Hence

$$\mathbf{H}_{\leftarrow} = \mathrm{Diag}(\mathbf{H}_1, \ldots, \mathbf{H}_H) \quad \text{and} \quad \mathbf{H}_{\rightarrow} = \mathrm{Diag}(\tilde{\mathbf{H}}_1, \ldots, \tilde{\mathbf{H}}_H).$$

In addition

$$\mathbf{H}_{\uparrow} = \left[\mathbf{0}_{N,W}, \mathbf{H}_{\uparrow}^1\right] \quad \text{and} \quad \mathbf{H}_{\downarrow} = \left[\mathbf{H}_{\downarrow}^1, \mathbf{0}_{N,W}\right]$$

with

$$\mathbf{H}_{\uparrow}^1 = \begin{pmatrix} & & -1 & 0 & \ldots & 0 \\ & W & & \ddots & \ddots & \vdots \\ & & 1 & & \ddots & 0 \\ & & 0 & \ddots & & -1 \\ N-W & & \vdots & \ddots & \ddots & \\ & & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times(N-W)}$$

$\mathbf{H}_{\downarrow}^1 = -\mathbf{H}_{\uparrow}^1$.

The only terms in $\frac{1}{2}\|\mathbf{AH}\|_{\mathrm{F}}^2$ related to $\mathbf{a}_n$ are

$$\begin{aligned} \phi(\mathbf{a}_n) =& \underbrace{\frac{1}{2}\left(\sum_{k=0}^{3} h_{n,n+kN}^2\right)}_{cA_n}\|\mathbf{a}_n\|_2^2 \\ &+ \underbrace{\left(\sum_{\substack{i=1\\ i\neq n}}^{N}\sum_{k=0}^{3} h_{n,n+kN}h_{i,n+kN}\mathbf{a}_i^T\right)}_{\mathbf{c}_n^T}\mathbf{a}_n. \end{aligned} \tag{32}$$

### B. Endmember penalization

*1) Distance between the endmembers and reference signatures:* The distance between the endmembers and the available reference signatures is

$$\Psi(\mathbf{M}) = \frac{1}{2}\left\|\mathbf{M}-\mathbf{M}_0\right\|_{\mathrm{F}}^2 = \frac{1}{2}\sum_{\ell=1}^{L}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{m}}_{0,\ell}\right\|_2^2. \quad (33)$$

As a consequence, the penalty for the $\ell$th band is

$$\psi(\widetilde{\mathbf{m}}_\ell) = \frac{1}{2}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{m}}_{0,\ell}\right\|_2^2. \quad (34)$$

*2) Mutual distance between the endmembers:* The distance between the different endmembers can be expressed as follows

$$\begin{aligned}\Psi(\mathbf{M}) &= \frac{1}{2}\sum_{i=1}^{K}\left(\sum_{\substack{j=1\\ j\neq i}}^{K}\left\|\mathbf{m}_i-\mathbf{m}_j\right\|_2^2\right)\\ &= \frac{1}{2}\sum_{k=1}^{K}\left\|\mathbf{M}\mathbf{G}_k\right\|_{\mathrm{F}}^2 = \frac{1}{2}\left\|\mathbf{M}\mathbf{G}\right\|_{\mathrm{F}}^2\end{aligned} \quad (35)$$

with

$$\mathbf{G} = \left[\begin{array}{c|c|c|c} \mathbf{G}_1 & \cdots & \mathbf{G}_K \end{array}\right] \in \mathbb{R}^{K\times K^2}$$

and for $k = 1,\ldots,K$

$$\mathbf{G}_k = -\mathbf{I}_K + \mathbf{e}_k\mathbf{1}_K^T$$

where $\mathbf{e}_k$ denotes the $k$th vector in the canonical basis of $\mathbb{R}^K$. Hence

$$\psi(\widetilde{\mathbf{m}}_\ell) = \frac{1}{2}\sum_{k=1}^{K}\left\|\widetilde{\mathbf{m}}_\ell\mathbf{G}_k\right\|_2^2. \quad (36)$$

*3) Volume and endmember positivity constraint:* The volume penalization is expressed using $\mathbf{T}$, hence the need to find a condition equivalent to the positivity of $\mathbf{M}$ (see [36]). We will first analyze the general expression of the volume penalization with respect to $\widetilde{\mathbf{t}}_k$, and then give a condition on $\mathbf{T}$ ensuring the positivity of $\mathbf{M}$ (respectively $\mathbf{M}+\mathbf{dM}_n$ when endmember variability is considered).

*a) Volume:* The determinant of a matrix $\mathbf{X} \in \mathbb{R}^{K\times K}$ can be developed along its $i$th row yielding

$$\det(\mathbf{X}) = \sum_j (-1)^{i+j}x_{ij}\det(\mathbf{X_{ij}}) = \widetilde{\mathbf{x}}_i\mathbf{f}_i$$

with

$$\mathbf{f}_i = \left[(-1)^{i+j}\det(\mathbf{X_{ij}})\right]_{j=1}^{K} \in \mathbb{R}^K.$$

Consequently, for $k = 1,\ldots,K-1$

$$\det\begin{pmatrix}\mathbf{T}\\ \mathbf{1}_K^T\end{pmatrix} = \widetilde{\mathbf{t}}_k\mathbf{f}_k.$$

Using previous developments

$$\psi(\widetilde{\mathbf{t}}_k) = \frac{1}{2(K-1)!^2}(\widetilde{\mathbf{t}}_k\mathbf{f}_k)^2. \quad (37)$$

*b) Positivity constraint on $\boldsymbol{M}$:* Using the following notations

$$\begin{aligned}\mathbf{Y} &= \mathbf{U}\mathbf{Y}_{\text{proj.}} + \bar{\mathbf{Y}}_1, & \bar{\mathbf{Y}}_1 &= [\bar{\mathbf{y}}|\ldots|\bar{\mathbf{y}}] \in \mathbb{R}^{L\times N}\\ \mathbf{M} &= \mathbf{U}\mathbf{T} + \bar{\mathbf{Y}}_2, & \bar{\mathbf{Y}}_2 &= [\bar{\mathbf{y}}|\ldots|\bar{\mathbf{y}}] \in \mathbb{R}^{L\times K}\end{aligned}$$

one has

$$m_{\ell r} = \sum_j u_{\ell j}t_{jr} + \bar{y}_\ell = \sum_{j\neq k} u_{\ell j}t_{jr} + u_{\ell k}t_{kr} + \bar{y}_\ell.$$

The positivity constraint for $m_{\ell r}$ can then be expressed as

$$t_{kr} \geq -\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}}{u_{\ell k}}.$$

Introducing the two sets of integers

$$\begin{aligned}\mathcal{U}_k^+ &= \{\ell | u_{\ell k} > 0\}\\ \mathcal{U}_k^- &= \{\ell | u_{\ell k} < 0\}\end{aligned}$$

the previous equation implies that $t_{kr} \in [t_{kr}^-, t_{kr}^+]$, with

$$t_{kr}^- = \max_{\ell\in\mathcal{U}_k^+}\left(-\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}}{u_{\ell k}}\right) \quad (38)$$

$$t_{kr}^+ = \min_{\ell\in\mathcal{U}_k^-}\left(-\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}}{u_{\ell k}}\right). \quad (39)$$

*4) Positivity constraint on $\boldsymbol{M}$ and $\boldsymbol{dM}$:* This case differs from the previous one as the positivity constraint must be verified simultaneously by $\mathbf{M}$ and $\mathbf{M}_n = \mathbf{M} + \mathbf{dM}_n$. We will consequently derive a condition similar to (38). Let $\mathbf{T}_n$ be the projection of $\mathbf{M}_n$ in the PCA subspace

$$\mathbf{M}_n = \mathbf{U}\mathbf{T}_n + \bar{\mathbf{Y}}_2.$$

Since

$$\mathbf{T}_n = \mathbf{V}\left(\mathbf{M}_n - \bar{\mathbf{Y}}_2\right) = (\mathbf{T} + \mathbf{dT}_n) + \underbrace{\mathbf{V}\bar{\mathbf{Y}}_2}_{\mathbf{z}}$$

with

$$\left\{\begin{array}{ll}\mathbf{T} & = \mathbf{V}\left(\mathbf{M} - \bar{\mathbf{Y}}_2\right)\\ \mathbf{dT}_n & = \mathbf{V}\left(\mathbf{dM}_n - \bar{\mathbf{Y}}_2\right)\end{array}\right.$$

the positivity constraint can be written

$$\begin{aligned} m_{\ell r}^n \geq 0 \quad &\Leftrightarrow \quad t_{kr}^n \geq -\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}^n}{u_{\ell k}}\\ &\Leftrightarrow \quad t_{kr}^n \in \left[t_{kr}^{n-}, t_{kr}^{n+}\right]\end{aligned}$$

with

$$t_{kr}^n = t_{kr} + dt_{kr}^n + z_{kr} \quad (40)$$

$$t_{kr}^{n-} = \max_{\ell\in\mathcal{U}_k^+}\left(-\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}^n}{u_{\ell k}}\right) \quad (41)$$

$$t_{kr}^{n+} = \min_{\ell\in\mathcal{U}_k^-}\left(-\frac{\bar{y}_\ell + \sum_{j\neq k} u_{\ell j}t_{jr}^n}{u_{\ell k}}\right). \quad (42)$$

We introduce the functions $g_k$ defined by

$$g_k: \begin{array}{ccc} \mathbb{R}^{1\times K} & \to & \mathbb{R}^{2(N+1)\times K} \\ \widetilde{\mathbf{x}} & \mapsto & \begin{bmatrix} \widetilde{\mathbf{x}} - \widetilde{\mathbf{t}}_k^- \\ -\widetilde{\mathbf{x}} + \widetilde{\mathbf{t}}_k^+ \\ \mathbf{1}_N(\widetilde{\mathbf{x}} + \widetilde{\mathbf{z}}_k) + \begin{pmatrix} \widetilde{\mathbf{dt}}_{1,k} - \widetilde{\mathbf{t}}_{1,k}^- \\ \vdots \\ \widetilde{\mathbf{dt}}_{N,k} - \widetilde{\mathbf{t}}_{N,k}^- \end{pmatrix} \\ -\mathbf{1}_N(\widetilde{\mathbf{x}} + \widetilde{\mathbf{z}}_k) + \begin{pmatrix} \widetilde{\mathbf{t}}_{1,k}^+ - \widetilde{\mathbf{dt}}_{1,k} \\ \vdots \\ \widetilde{\mathbf{t}}_{N,k}^+ - \widetilde{\mathbf{dt}}_{N,k} \end{pmatrix} \end{bmatrix}, \end{array} \tag{43}$$

where

$$\begin{aligned} \widetilde{\mathbf{t}}_k^+ &= [t_{k1}^+, \cdots, t_{kK}^+] \\ \widetilde{\mathbf{t}}_k^- &= [t_{k1}^-, \cdots, t_{kK}^-]. \end{aligned}$$

Finally the positivity constraint on the sum of the endmembers and their variability can be written

$$\widetilde{\mathbf{m}}_\ell + \widetilde{\mathbf{dm}}_\ell^n \succeq \mathbf{0}_K^T \quad \forall \ell, \quad \forall n \tag{44}$$

$$\Leftrightarrow g_k(\widetilde{\mathbf{t}}_k) \succeq \mathbf{0}_{2(N+1),K} \quad \forall k = 1, \ldots, K-1. \tag{45}$$

### C. Variability penalization

The variability energy penalty is

$$\Upsilon(\mathbf{dM}) = \frac{1}{2}\|\mathbf{dM}\|_{\mathrm{F}}^2 \Rightarrow \upsilon(\mathbf{dM}_n) = \frac{1}{2}\|\mathbf{dM}_n\|_{\mathrm{F}}^2. \tag{46}$$

## Appendix B
## Solutions to the optimization sub-problems

### A. Resolution with respect to A

Using (32), the scaled augmented Lagrangian (23) becomes

$$\begin{aligned} \mathcal{L}_{\mu_n^{(\mathbf{A})}}\left(\mathbf{a}_n, \mathbf{w}_n^{(\mathbf{A})}, \boldsymbol{\lambda}_n^{(\mathbf{A})}\right) &= \frac{1}{2}\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\|_2^2 \\ &+ \frac{\alpha}{2}\left(cA_n\|\mathbf{a}_n\|_2^2 + 2\mathbf{c}_n^T\mathbf{a}_n\right) + \mathcal{I}_{\mathcal{S}_K^+}\left(\mathbf{w}_n^{(\mathbf{A})}\right) \\ &+ \frac{\mu_n^{(\mathbf{A})}}{2}\left\|\mathbf{Q}\mathbf{a}_n + \mathbf{R}\mathbf{w}_n^{(\mathbf{A})} - \mathbf{s} + \boldsymbol{\lambda}_n^{(\mathbf{A})}\right\|_2^2. \end{aligned}$$

Thus, for $n = 1, \ldots, N$

$$\begin{aligned} \mathbf{a}_n^* = &\left[(\mathbf{M} + \mathbf{dM}_n)^T(\mathbf{M} + \mathbf{dM}_n) + \mu_n^{(\mathbf{A})}\mathbf{Q}^T\mathbf{Q} + \alpha cA_n\mathbf{I}_K\right]^{-1} \\ &\left[(\mathbf{M} + \mathbf{dM}_n)^T\mathbf{y}_n - \alpha\mathbf{c}_n + \mu_n^{(\mathbf{A})}\mathbf{Q}^T\left(\mathbf{s} - \mathbf{R}\mathbf{w}_n^{(\mathbf{A})} - \boldsymbol{\lambda}_n^{(\mathbf{A})}\right)\right] \end{aligned} \tag{47}$$

and

$$\mathbf{w}_n^{(\mathbf{A})^*} = \max\left(\mathbf{a}_n + \boldsymbol{\lambda}_{n,1:K}^{(\mathbf{A})}, \mathbf{0}_K\right) \tag{48}$$

where $\boldsymbol{\lambda}_{n,1:K}^{(\mathbf{A})}$ is the vector composed of the $K$ first elements of $\boldsymbol{\lambda}_n^{(\mathbf{A})}$ and the max must be understood as a term-wise operator. In the absence of any penalization, the solution is obtained by making $\alpha = 0$ in the previous equations.

### B. Resolution with respect to M

*1) Distance between the endmembers and reference spectral signatures:* Using (34), the scaled augmented Lagrangian (26) is

$$\begin{aligned} \mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \mathbf{W}_\ell^{(\mathbf{M})}, \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right) &= \frac{1}{2}\left\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\right\|_2^2 \\ &+ \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\mathbf{e}\widetilde{\mathbf{m}}_\ell - \mathbf{W}_\ell^{(\mathbf{M})} + \mathbf{F}_\ell + \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right\|_{\mathrm{F}}^2 \\ &+ \frac{\beta}{2}\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{m}}_{\ell,0}\|_2^2 + \mathcal{I}_{\mathcal{S}^+}\left(\mathbf{W}_\ell^{(\mathbf{M})}\right). \end{aligned}$$

Thus

$$\begin{aligned} \widetilde{\mathbf{m}}_\ell^* = &\left[\left(\widetilde{\mathbf{y}}_\ell - \widetilde{\boldsymbol{\delta}}_\ell\right)\mathbf{A}^T + \beta\widetilde{\mathbf{m}}_{\ell,0} + \right. \\ &\left. \mu_\ell^{(\mathbf{M})}\mathbf{e}^T\left(\mathbf{W}_\ell^{(\mathbf{M})} - \mathbf{F}_\ell - \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right)\right] \\ &\left[\mathbf{A}\mathbf{A}^T + \mu_\ell^{(\mathbf{M})}\left(\mathbf{e}^T\mathbf{e} + \beta\right)\mathbf{I}_K\right]^{-1} \end{aligned} \tag{49}$$

and for $k = 1, \ldots, K$

$$\mathbf{w}_{\ell,k}^{(\mathbf{M})*} = \max\left([\mathbf{e}\widetilde{\mathbf{m}}_\ell + \mathbf{F}_\ell + \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}]_k, \mathbf{0}_{N+1}\right). \tag{50}$$

In the absence of any endmember penalization, the solution is obtained by making $\beta = 0$ in the previous equation.

*2) Mutual distance between the endmembers:* Using (36), the scaled augmented Lagrangian (26) is

$$\begin{aligned} \mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \mathbf{W}_\ell^{(\mathbf{M})}, \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right) &= \frac{1}{2}\left\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\right\|_2^2 \\ &+ \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\mathbf{e}\widetilde{\mathbf{m}}_\ell - \mathbf{W}_\ell^{(\mathbf{M})} + \mathbf{F}_\ell + \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right\|_{\mathrm{F}}^2 \\ &+ \frac{\beta}{2}\sum_{k=1}^K\|\widetilde{\mathbf{m}}_\ell\mathbf{G}_k\|_2^2 + \mathcal{I}_{\mathcal{S}^+}\left(\mathbf{W}_\ell^{(\mathbf{M})}\right). \end{aligned}$$

Thus

$$\begin{aligned} \widetilde{\mathbf{m}}_\ell^* = &\left[\left(\widetilde{\mathbf{y}}_\ell - \widetilde{\boldsymbol{\delta}}_\ell\right)\mathbf{A}^T + \mu_\ell^{(\mathbf{M})}\mathbf{e}^T\left(\mathbf{W}_\ell^{(\mathbf{M})} - \mathbf{F}_\ell - \boldsymbol{\Lambda}_\ell^{(\mathbf{M})}\right)\right] \\ &\left[\mathbf{A}\mathbf{A}^T + \beta\sum_{k=1}^K\mathbf{G}_k\mathbf{G}_k^T + \mu_\ell^{(\mathbf{M})}\left(\mathbf{e}^T\mathbf{e}\right)\mathbf{I}_K\right]^{-1} \end{aligned} \tag{51}$$

with $\mathbf{W}_\ell^{(\mathbf{M})}$ given by (50).

*3) Volume penalization:* Since the penalty is expressed with respect to the variable $\mathbf{T}$, the optimization sub-problems related to the endmembers have to be re-written accordingly. Using the notations

$$\boldsymbol{\Delta}_T = \left[\ \mathbf{dT}_1\mathbf{a}_1 \ \middle|\ \ldots \ \middle|\ \mathbf{dT}_N\mathbf{a}_N\ \right] \tag{52}$$

$$\bar{\mathbf{Y}}_1 = [\bar{\mathbf{y}}|\cdots|\bar{\mathbf{y}}] \in \mathbb{R}^{L\times N} \tag{53}$$

$$\bar{\mathbf{Y}}_2 = [\bar{\mathbf{y}}|\cdots|\bar{\mathbf{y}}] \in \mathbb{R}^{L\times K} \tag{54}$$

we obtain

$$\begin{aligned} \|\mathbf{Y} - \mathbf{M}\mathbf{A} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2 &= \left\|\mathbf{U}\left(\mathbf{Y}_{\text{proj.}} - \mathbf{T}\mathbf{A} - \boldsymbol{\Delta}_T\right) + \bar{\mathbf{Y}}_1 - 2\bar{\mathbf{Y}}_2\mathbf{A}\right\|_{\mathrm{F}}^2 \\ &= \|\mathbf{U}\left(\mathbf{Y}_{\text{proj.}} - \mathbf{T}\mathbf{A} - \boldsymbol{\Delta}_T\right)\|_{\mathrm{F}}^2 \\ &+ 2\left\langle\mathbf{U}\left(\mathbf{Y}_{\text{proj.}} - \mathbf{T}\mathbf{A} - \boldsymbol{\Delta}_T\right)\middle|\bar{\mathbf{Y}}_1 - 2\bar{\mathbf{Y}}_2\mathbf{A}\right\rangle \\ &+ \left\|\bar{\mathbf{Y}}_1 - 2\bar{\mathbf{Y}}_2\mathbf{A}\right\|_{\mathrm{F}}^2. \end{aligned}$$

The only terms depending on $\mathbf{T}$ are

$$\|\mathbf{Y}_{\text{proj.}} - \mathbf{TA}\|_{\text{F}}^2 + 2\Big\langle \underbrace{\boldsymbol{\Delta}_T - \mathbf{U}^T(\bar{\mathbf{Y}}_1 - 2\bar{\mathbf{Y}}_2\mathbf{A})}_{\mathbf{S}} \Big| \mathbf{TA} \Big\rangle$$

with

$$\langle \mathbf{S}|\mathbf{TA}\rangle = \text{Tr}\left(\mathbf{S}^T\mathbf{TA}\right) = \sum_{n=1}^{N}\left(\sum_{j=1}^{K-1} s_{jn}\widetilde{\mathbf{t}}_j\mathbf{a}_n\right).$$

For $k = 1, \ldots, K-1$, the resulting sub-problems are

$$\widetilde{\mathbf{t}}_k^* = \arg\min_{\widetilde{\mathbf{t}}_k} \left\{ \begin{array}{c} \frac{1}{2}\left\|\widetilde{\mathbf{y}}_k^{\text{proj.}} - \widetilde{\mathbf{t}}_k\mathbf{A}\right\|_2^2 + \sum_{n=1}^{N}(s_{kn}\widetilde{\mathbf{t}}_k\mathbf{a}_n) \\ +\frac{\beta}{2(K-1)!^2}(\widetilde{\mathbf{t}}_k\mathbf{f}_k)^2 \\ \text{s.t.} \quad g_k(\widetilde{\mathbf{t}}_k) \succeq \mathbf{0}_{2(N+1),K} \end{array} \right\}.$$

Introduce the splitting variables $\mathbf{W}_k^{(\mathbf{T})}$ such that

$$g_k(\widetilde{\mathbf{t}}_k) = \mathbf{W}_k^{(\mathbf{T})} \quad \forall k = 1, \ldots, K-1. \tag{55}$$

According to (37), the scaled augmented Lagrangian is

$$\begin{aligned} &\mathcal{L}_{\mu_k^{(\mathbf{T})}}(\widetilde{\mathbf{t}}_k, \mathbf{W}_k^{(\mathbf{T})}, \boldsymbol{\Lambda}_k^{(\mathbf{T})}) = \\ &\quad \frac{1}{2}\left\|\widetilde{\mathbf{y}}_k^{\text{proj.}} - \widetilde{\mathbf{t}}_k\mathbf{A}\right\|_2^2 + \sum_{n=1}^{N}\left(s_{nk}\widetilde{\mathbf{t}}_k\mathbf{a}_n\right) \\ &\quad + \frac{\beta}{2(K-1)!^2}\left(\widetilde{\mathbf{t}}_k\mathbf{f}_k\right)^2 + \mathcal{I}_{\mathcal{S}^+}\left(\mathbf{W}_k^{(\mathbf{T})}\right) \\ &\quad + \frac{\mu_k^{(\mathbf{T})}}{2}\left\|g_k(\widetilde{\mathbf{t}}_k) - \mathbf{W}_k^{(\mathbf{T})} + \boldsymbol{\Lambda}_k^{(\mathbf{T})}\right\|_{\text{F}}^2. \end{aligned} \tag{56}$$

Finally,

$$\begin{aligned} \widetilde{\mathbf{t}}_k^* = &\Bigg[\left(\widetilde{\mathbf{y}}_k^{\text{proj.}} - \widetilde{\mathbf{s}}_k\right)\mathbf{A}^T - 2\mu_k^{(\mathbf{T})}\left(N\widetilde{\mathbf{z}}_k + \sum_{n=1}^{N}\widetilde{\mathbf{dt}}_{n,k}\right) \\ &+\mu_k^{(\mathbf{T})}\left(\widetilde{\mathbf{t}}_k^- + \widetilde{\mathbf{t}}_k^+ + \sum_{n=1}^{N}(\widetilde{\mathbf{t}}_{n,k}^- + \widetilde{\mathbf{t}}_{n,k}^+)\right) \\ &+\mu_k^{(\mathbf{T})}\left[1 \quad -1 \quad \mathbf{1}_N^T \quad -\mathbf{1}_N^T\right]\left(\mathbf{W}_k^{(\mathbf{T})} - \boldsymbol{\Lambda}_k^{(\mathbf{T})}\right)\Bigg] \\ &\left[\mathbf{AA}^T + \frac{\beta}{(K-1)!^2}\mathbf{f}_k\mathbf{f}_k^T + 2(N+1)\mu_k^{(\mathbf{T})}\mathbf{I}_K\right]^{-1} \end{aligned} \tag{57}$$

where $\mathbf{Z} = \mathbf{V}\bar{\mathbf{Y}}_2$ and for $p = 1, \ldots, K$

$$\mathbf{w}_{k,p}^{(\mathbf{T})*} = \max\left(\left[g_k(\widetilde{\mathbf{t}}_k) + \boldsymbol{\Lambda}_k^{(\mathbf{T})}\right]_p, \mathbf{0}_{2(N+1)}\right). \tag{58}$$

### C. Resolution with respect to dM

Using (46), the scaled augmented Lagrangian (28) is

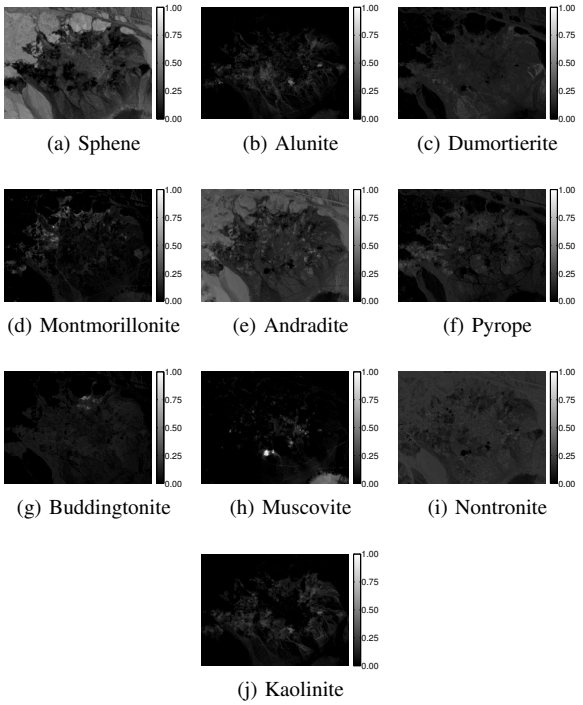
$$\begin{aligned} &\mathcal{L}_{\mu_n^{(\mathbf{dM})}}\left(\mathbf{dM}_n, \mathbf{W}_n^{(\mathbf{dM})}, \boldsymbol{\Lambda}_{\mathbf{dM_n}}\right) = \\ &\quad \frac{1}{2}\left\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a_n}\right\|_2^2 \\ &\quad + \frac{\mu_n^{(\mathbf{dM})}}{2}\left\|\mathbf{dM}_n + \mathbf{M} - \mathbf{W}_n^{(\mathbf{dM})} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right\|_{\text{F}}^2 \\ &\quad + \frac{\gamma}{2}\left\|\mathbf{dM}_n\right\|_{\text{F}}^2 + \mathcal{I}_{\mathcal{S}^+}\left(\mathbf{W}_n^{(\mathbf{dM})}\right). \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{dM}_n^* = &\Big[(\mathbf{y_n} - \mathbf{Ma}_n)\mathbf{a}_n^T + \mu_n^{(\mathbf{dM})}\left(\mathbf{W}_n^{(\mathbf{dM})} - \mathbf{M} - \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right)\Big] \\ &\left[\mathbf{a}_n\mathbf{a}_n^T + (\mu_n^{(\mathbf{dM})} + \gamma)\mathbf{I}_K\right]^{-1} \end{aligned} \tag{59}$$

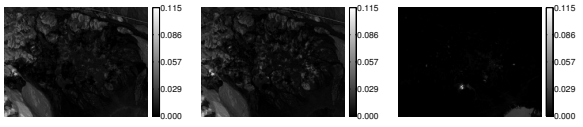
and for $p = 1, \ldots, K$

$$\mathbf{w}_{n,p}^{(\mathbf{dM})*} = \max\left(\left[\mathbf{dM}_n + \mathbf{M} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right]_p, \mathbf{0}_K\right).$$



(a) Sphene (b) Alunite (c) Dumortierite (d) Montmorillonite (e) Andradite (f) Pyrope (g) Buddingtonite (h) Muscovite (i) Nontronite (j) Kaolinite

Fig. 7. ssvcaBCD/ADMM abundance results (Cuprite scene). The given identification is based on a visual comparison with the results obtained in [28].



(a) Sphene (b) Andradite (c) Muscovite

Fig. 8. Spatial distribution of the variability with respect to the endmembers presenting the most significant level of variability. The variability is presented in terms of energy ($\frac{1}{\sqrt{L}}\|\mathbf{dm}_{n,k}\|_2$ for the $k$th endmember in the $n$th pixel) for visualization purpose.

## REFERENCES


[1] J. M. 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 J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

[2] A. Zare and K. C. Ho, "Endmember variability in hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

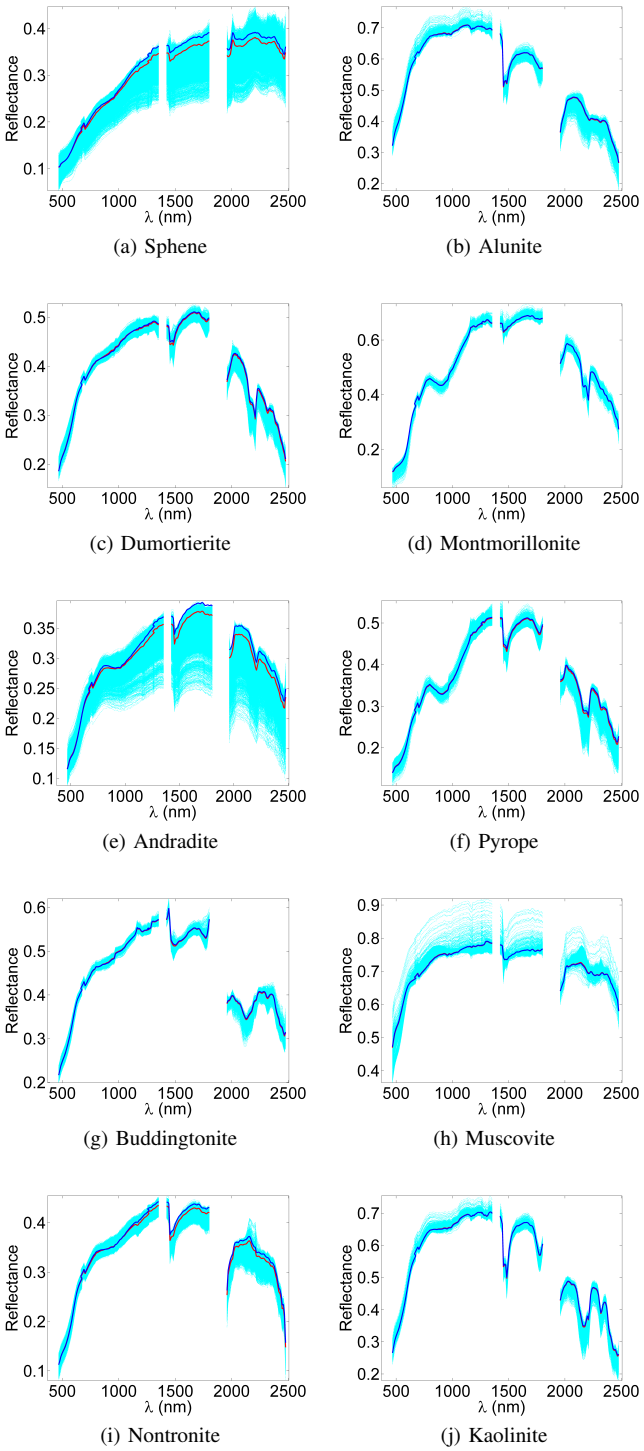



Fig. 9. ssvcaBCD/ADMM endmembers (Cuprite scene). The BCD/ADMM-estimated endmembers (red lines) are given with typical examples of the estimated variability (cyan dotted lines), and the given identification is based on a visual comparison with the results obtained in [28]. The endmembers estimated by the VCA algorithm are given in blue lines for comparison.


[3] 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 Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, June 2010.

[4] X. Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 1994–2003, June 2014.

[5] B. Somers, M. Zortea, A. Plaza, and G. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 2, pp. 396–408, April 2012.

[6] J. Jin, B. Wang, and L. Zhang, "A novel approach based on Fisher discriminant null space for decomposition of mixed pixels in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 699–703, Oct. 2010.

[7] M. A. Veganzones, L. Drumetz, G. Tochon, M. D. Mura, A. Plaza, J. M. Bioucas-Dias, and J. Chanussot, "A new extended linear mixing model to adress spectral variability," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Lausanne, Switzerland, June 2014.

[8] E. C. Johnson and D. L. Jones, "Joint recovery of sparse signals and parameter perturbations with parameterized measurement models," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 5900–5904.

[9] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Grenoble, France, Aug. 2009.

[10] M. V. Alfonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sept. 2010.

[11] ——, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, March 2011.

[12] M. S. C. Almeida and M. A. T. Figueiredo, "Blind image deblurring with unknown boundaries using the alternating direction method of multipliers," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Melbourne, Australia, 2013, pp. 586–590.

[13] ——, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *arXiv:1210.2687v2*, March 2013.

[14] A. Matakos, S. Ramani, and J. A. Fessler, "Accelerated edge-preserving image restoration with boundary artifacts," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2019–2029, May 2013.

[15] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "GLUP: yet another algorithm for blind unmixing of hyperspectral data," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Lausanne, Switzerland, June 2014, pp. 1–4.

[16] ——, "Blind and fully constrained unmixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5510–5518, Dec. 2014.

[17] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[18] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J.-Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using an eigen-gap approach," 2015, submitted. [Online]. Available: http://arxiv.org/pdf/1501.05552v1.pdf

[19] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundance in hyperspectral images with $\ell_1$-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[20] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–776, March 2007.

[21] T.-H. Chan, Wing-KinMa, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.

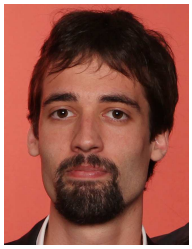
[22] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[23] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using Bayesian nonnegative matrix factorization with volume prior," *J. Signal Process. Sys.*, vol. 65, no. 3, pp. 479–496, Nov. 2011.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[25] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[26] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 1-2, no. 146, pp. 459–494, July 2013.
[27] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Estimation de variabilité pour le démélange non-supervisé d'images hyperspectrales," in *Actes du XXVième Colloque GRETSI*, Lyon, France, Sept. 2015, in French, to appear.
[28] J. M. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, April 2005.
[29] D. C. Heinz and C.-I Chang, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, pp. 1144–1159, May 2000.
[30] D. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping chaparral in the Santa Monica mountain using multiple endmember spectral mixture models," *Remote Sens. Environment*, vol. 65, no. 3, pp. 267–279, Sept. 1998.
[31] M. A. Goenaga, M. C. Torres-Madronero, M. Velez-Reyez, S. J. V. Bloem, and J. D. Chinea, "Unmixing analysis of a time series of hyperion images over the Guánica dry forest in Puerto Rico," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 6, no. 2, pp. 329–338, April 2013.
[32] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Bilinear models for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Lisbon, Portugal, June 2011, pp. 1–4.
[33] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, March 2015, to appear. [Online]. Available: http://arxiv.org/abs/1401.5649/
[34] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.
[35] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 89–94, Jan. 2014.
[36] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.
[37] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *arXiv:0912.3522v4*, May 2010.
[38] D. C. Heinz and C. -I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, March 2001.
[39] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.



**Pierre-Antoine Thouvenin** received the state engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse (INP Toulouse), both in 2014. He is currently working toward the Ph.D. degree within the Signal and Communications Group of the IRIT Laboratory, Toulouse, France. His current research interests include hyperspectral unmixing and optimization techniques.

**Nicolas Dobigeon** (S'05–SM'08–SM'13) received the state engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse (INP Toulouse), both in June 2004, as well as the Ph.D. degree and Habilitation à Diriger des Recherches in Signal Processing from the INP Toulouse in 2007 and 2012, respectively. He was a Post-Doctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA, from 2007 to 2008.

Since 2008, he has been with the National Polytechnic Institute of Toulouse (INP-ENSEEIHT, University of Toulouse) where he is currently an Associate Professor. He conducts his research within the Signal and Communications Group of the IRIT Laboratory and he is also an affiliated faculty member of the Telecommunications for Space and Aeronautics (TeSA) cooperative laboratory. His current research interests include statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing, biomedical imaging and genomics.

**Jean-Yves TOURNERET** (SM'08) received the ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been involved in the organization of several conferences including the European conference on signal processing EUSIPCO'02 (program chair), the international conference ICASSP'06 (plenaries), the statistical signal processing workshop SSP'12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangements), the statistical signal processing workshop SSP'2014 (special sessions), the workshop on machine learning for signal processing MLSP'2014 (special sessions). He has been the general chair of the CIMI workshop on optimization and statistics in image processing hold in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and of the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001-2007, 2010-present). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008-2011, 2015-present) and for the EURASIP journal on Signal Processing (2013-present).