Online Unmixing of Multitemporal Hyperspectral Images accounting for Spectral Variability

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

Abstract—Hyperspectral unmixing is aimed at identifying the reference spectral signatures composing an hyperspectral image and their relative abundance fractions in each pixel. In practice, the identified signatures may vary spectrally from an image to another due to varying acquisition conditions, thus inducing possibly significant estimation errors. Against this background, hyperspectral unmixing of several images acquired over the same area is of considerable interest. Indeed, such an analysis enables the endmembers of the scene to be tracked and the corresponding endmember variability to be characterized. Sequential endmember estimation from a set of hyperspectral images is expected to provide improved performance when compared to methods analyzing the images independently. However, the significant size of hyperspectral data precludes the use of batch procedures to jointly estimate the mixture parameters of a sequence of hyperspectral images. Provided that each elementary component is present in at least one image of the sequence, we propose to perform an online hyperspectral unmixing accounting for temporal endmember variability. The online hyperspectral unmixing is formulated as a two-stage stochastic program, which can be solved using a stochastic approximation. The performance of the proposed method is evaluated on synthetic and real data. A comparison with independent unmixing algorithms finally illustrates the interest of the proposed strategy.

Index Terms—Hyperspectral imagery, perturbed linear unmixing (PLMM), endmember temporal variability, two-stage stochastic program, stochastic approximation (SA).

I. INTRODUCTION

HYPERSPECTRAL imagery has known an increasing interest over the past decades due to the significant spectral information it conveys. Acquired in hundreds of contiguous spectral bands (e.g., from $300\,\text{nm}$ to $2600\,\text{nm}$ for the AVIRIS sensor), hyperspectral (HS) images facilitate the identification of the elements composing the imaged scene[1]. However, the high spectral resolution of these images is mitigated by their lower spatial resolution, which results in pixel spectra composed of mixtures of reference signatures. Spectral unmixing consists of determining the reference spectral signatures composing the data – referred to as *endmembers* – and their abundance fractions in each pixel according to a predefined mixture model accounting for several environmental factors (declivity, multiple reflections, ...). Provided

[1]Note that some airborne sensors cover larger wavelength range, while some mounted sensors can offer millimeter spatial resolutions.

microscopic interactions between the materials of the imaged scene are negligible and the relief of the scene is flat, a linear mixing model (LMM) is traditionally used to describe the data [1]. However, varying acquisition conditions such as illumination or natural evolution of the scene may significantly alter the shape and the amplitude of the spectral signatures acquired, thus affecting the extracted endmembers from an image to another. In this context, HS unmixing of several images acquired over the same area at different time instants can be of considerable interest. Indeed, such an analysis enables the endmembers of the scene and endmember variability to be assessed, thus improving endmember estimation when compared to independent image analyses performed with any state-of-the-art unmixing method.

So far, spatial variability within a given image has been considered in various models either derived from a statistical or a deterministic point of view [2]. The first class of methods assumes that the endmember spectra are realizations of multivariate distributions [3]–[5]. The second class of methods represents endmember signatures as members of spectral libraries associated with each material (bundles) [6]. Another recently proposed approach consists in estimating the parameters of an explicit variability model [7]. To the best of our knowledge, spatio-temporal variability has been analyzed for the first time in the Bayesian framework proposed in [8]. Another recent contribution similarly resorts to a batch estimation technique to address spectral unmixing of multi-temporal HS images [9]. However, HS unmixing using a significant number of images or several large images precludes the use of batch estimation procedures as in [8], [9] due to limited memory and computational resources. Since online estimation procedures enable data to be sequentially incorporated into the estimation process without the need to simultaneously load all the data into memory, we focus in this paper on the design of an online HS unmixing method accounting for temporal variability.

Since the identified endmembers can be considered as time-varying instances of reference endmembers, we use the perturbed linear mixing model (PLMM) proposed in [7] to account for spectral variability. However, inspired by the works presented in [10], [11], we formulate the unmixing problem as a two-stage stochastic program that allows the model parameters to be estimated online contrary to the algorithm proposed in [7]. To the best of our knowledge, it is the first time HS unmixing accounting for temporal variability has been formulated as a two-stage stochastic program solved by an

TABLE I
NOTATIONS.

N	number of pixels
L	number of spectral bands
R	number of endmembers
T	number of images
$\mathbf{y}_{nt} \in \mathbb{R}^L$	nth pixel of the tth image
$\mathbf{Y}_t \in \mathbb{R}^{L \times N}$	lexicographically ordered pixels of the tth image
$\mathbf{M} \in \mathbb{R}^{L \times R}$	endmember matrix
$\mathbf{dM}_t \in \mathbb{R}^{L \times R}$	tth variability matrix
$\mathbf{A}_t \in \mathbb{R}^{R \times N}$	tth abundance matrix
$\mathbf{a}_{nt} \in \mathbb{R}^R$	nth column of the matrix $\mathbf{A}_t$
$\succeq$	component-wise inequality
$\mathcal{Y}$	$[0,1]^{L \times N}$
$\mathcal{M}$	$[0,1]^{L \times R}$
$\mathcal{S}_R$	unit simplex of $\mathbf{R}^R$
$\mathcal{A}_R$	$\left\{ \mathbf{A} \in \mathbb{R}^{R \times N} \big| \mathbf{a}_n \in \mathcal{S}_R, \forall n \in [\![1, N]\!] \right\}$
$\mathcal{B}_{\mathrm{F}}(\mathbf{Z}, \kappa)$	$\left\{ \mathbf{X} \in \mathbb{R}^{L \times R} \big| \|\mathbf{X} - \mathbf{Z}\|_{\mathrm{F}} \leq \kappa \right\}, \kappa > 0$
$\mathcal{D}$	$\mathcal{B}_{\mathrm{F}}(\mathbf{0}, \sigma) \cap \left\{ \mathbf{dM} \big| \|\mathbb{E}[\mathbf{dM}]\|_{\mathrm{F}} \leq \kappa \right\}$
$\mathcal{D}_t$	$\mathcal{B}_{\mathrm{F}}(\mathbf{0}, \sigma) \cap \left\{ \mathbf{dM} \big\| \left\|\sum_{i=1}^{t-1} \mathbf{dM}_i + \mathbf{dM}\right\|_{\mathrm{F}} \leq t\kappa \right\}$
$\mathcal{Z}_t$	$\mathcal{A}_R \times \mathcal{D}_t$
$\mathcal{Q}(\mathbf{Y}_t, \mathbf{M})$	$\left\{ (\mathbf{A}, \mathbf{dM}) \in \mathcal{Z}_t \big| \nabla_{(\mathbf{A}, \mathbf{dM})} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \mathbf{0} \right\}$
$\mathcal{P}_{\mathcal{S}}$	projector on the set $\mathcal{S}$
$\mathcal{P}_+$	projector on $\left\{ \mathbf{X} \in \mathbb{R}^{L \times R} \big| \mathbf{X} \succeq \mathbf{0}_{L,R} \right\}$
$\langle \mathbf{X}, \mathbf{Y} \rangle$	matrix inner product $\mathrm{Tr}(\mathbf{X}^T \mathbf{Y})$
$\iota_{\mathcal{S}}(x)$	$\begin{cases} 0 & \text{if } x \in \mathcal{S} \\ +\infty & \text{otherwise.} \end{cases}$

online[2] algorithm.

The paper is organized as follows. The proposed PLMM accounting for temporal variability is introduced in Section II. Section III describes an online algorithm to solve the resulting optimization problem. Experimental results obtained on synthetic and real data are reported in Sections IV and V respectively. The results obtained with the proposed algorithm are systematically compared to those obtained with the vertex component analysis / fully constrained least squares (VCA [12] / FCLS [13], [14]), SISAL [15] / FCLS, the $\ell_{1/2}$ non-negative matrix factorization (NMF) [16] and the BCD/ADMM algorithm of [7], each method being independently applied to each image of the sequence. Section VI finally concludes this work.

II. PROBLEM STATEMENT

A. Perturbed linear mixing model (PLMM)

We consider HS images acquired at T different time instants over the same scene, assuming that at most R endmembers are present in the resulting time series and that the images share these R common endmembers. Each endmember does not need to be present in each image, but at least in one image of the time series. Given an *a priori* known number of endmembers R, the PLMM consists in representing each pixel $\mathbf{y}_{nt}$ by a linear combination of the R endmembers – denoted by $\mathbf{m}_r$ – affected by a perturbation vector $\mathbf{dm}_{rt}$ accounting for temporal endmember variability. The proposed model considers the case where the variability essentially results from the evolution of the scene or from the global acquisition conditions from one image to another. As a first

---

[2]The terminology "online" is slightly abusive in our context since the time difference between two consecutive images can extend to several months.

---

approximation, the variability is assumed to be constant on each image. The resulting PLMM can thus be written

$$\mathbf{y}_{nt} = \sum_{r=1}^{R} a_{rnt} \left(\mathbf{m}_r + \mathbf{dm}_{rt} \right) + \mathbf{b}_{nt} \quad (1)$$

for $n = 1, \ldots, N$ and $t = 1, \ldots, T$, where $\mathbf{y}_{nt}$ denotes the nth image pixel at time t, $\mathbf{m}_r$ is the rth endmember, a_{rnt} is the proportion of the rth endmember in the nth pixel at time t, and $\mathbf{dm}_{rt}$ denotes the perturbation of the rth endmember at time t. Finally, $\mathbf{b}_{nt}$ models the noise resulting from the data acquisition and the modeling errors. In matrix form, the PLMM (1) can be written as

$$\mathbf{Y}_t = (\mathbf{M} + \mathbf{dM}_t)\mathbf{A}_t + \mathbf{B}_t \quad (2)$$

where $\mathbf{Y}_t = [\mathbf{y}_{1t}, \ldots, \mathbf{y}_{Nt}]$ is an $L \times N$ matrix containing the pixels of the tth image, $\mathbf{M}$ denotes an $L \times R$ matrix containing the endmembers, $\mathbf{A}_t$ is an $R \times N$ matrix composed of the abundance vectors $\mathbf{a}_{nt}$, $\mathbf{dM}_t$ is an $L \times R$ matrix whose columns are the perturbation vectors associated with the tth image, and $\mathbf{B}_t$ is an $L \times N$ matrix accounting for the noise at time instant t. The non-negativity and sum-to-one constraints usually considered to reflect physical considerations are

$$\mathbf{A}_t \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}_t^T \mathbf{1}_R = \mathbf{1}_N, \forall t = 1, \ldots, T$$
$$\mathbf{M} \succeq \mathbf{0}_{L,R} \quad (3)$$

where $\succeq$ denotes a component-wise inequality. We also consider the following assumptions on the inherent variability of the observed scenes

$$\|\mathbf{dM}_t\|_{\mathrm{F}}^2 \leq \sigma^2, \quad \text{for } t = 1, \ldots, T \quad (4)$$

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{dM}_t \right\|_{\mathrm{F}}^2 \leq \kappa^2 \quad (5)$$

where σ and κ are fixed positive constants, and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. These two constraints can be interpreted in terms of the feasible domain of $\mathbf{M}$ and $\mathbf{dM}_t$. Indeed, introducing the perturbed endmembers $\mathbf{M}_t \triangleq \mathbf{M} + \mathbf{dM}_t$, the constraint (4) can be reformulated as

$$\|\mathbf{dM}_t\|_{\mathrm{F}}^2 = \|\mathbf{M} - \mathbf{M}_t\|_{\mathrm{F}}^2 \leq \sigma^2 \Leftrightarrow \mathbf{M} \in \bigcap_{t=1}^{T} \mathcal{B}_{\mathrm{F}}(\mathbf{M}_t, \sigma)$$

where $\mathcal{B}_{\mathrm{F}}(\mathbf{M}_t, \sigma)$ is the ball of center $\mathbf{M}_t$ and of radius σ. This highlights the fact that the number of constraints imposed on the endmembers increases with T, i.e., the more images are processed, the more information can be extracted in terms of endmember signatures. On the other hand, (5) constrains the perturbed endmembers to be distributed around the true endmembers, i.e., the endmember signatures $\mathbf{M}$ should reflect the average behavior of the perturbed endmembers $\mathbf{M}_t$ in the sequence. In practice, setting σ to a reasonable value is desirable from a modeling point of view, since very large perturbations should probably be interpreted as outliers, thus leading to the removal of the corrupted elements from the unmixing process. Note however that the algorithm proposed in Section III-B is independent from any consideration on the values of σ^2 and κ^2.

Remark. In practice, HS unmixing is performed on reflectance data, hence $\mathbf{Y}_t \in [0,1]^{L \times N}$. The abundance sum-to-one and non-negativity constraints further imply

$\mathbf{M} \in [0,1]^{L \times R}$. In fact, the compactness of both the data support and the space associated with the endmember constraints – denoted by $\mathcal{Y}$ and $\mathcal{M}$ respectively – is crucial for the convergence result given in Paragraph III-C. In addition, the images $\mathbf{Y}_t$ can be assumed to be independent and identically distributed (i.i.d.) since these images have been acquired by possibly different sensors at different time instants.

B. Problem formulation

In order to design an online estimation algorithm, the model (1) combined with the constraints (3) can be used to formulate a two-stage stochastic program consisting in estimating the endmembers present in the image sequence. Since only the endmembers are supposed to be commonly shared by the different images, we propose to minimize a marginal cost function obtained by marginalizing an instantaneous cost function over the abundances and the variability terms, so that the resulting cost only depends on the endmembers. Assuming the expectations are well-defined, we consider the following optimization problem

$$\min_{\mathbf{M} \in \mathcal{M}} g(\mathbf{M}) = \mathbb{E}_{\mathbf{Y},\mathbf{A},\mathbf{dM}}\Big[f\big(\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{dM}\big)\Big] \qquad (6)$$

where $\mathcal{M} = [0,1]^{L \times R}$ and where the function f is defined as

$$f(\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \frac{1}{2}\|\mathbf{Y} - (\mathbf{M} + \mathbf{dM})\mathbf{A}\|_{\mathrm{F}}^2 \\ + \alpha \Phi(\mathbf{A}) + \beta \Psi(\mathbf{M}) + \gamma \Upsilon(\mathbf{dM}) \qquad (7)$$

Φ, Ψ and Υ denote appropriate penalization terms on the abundances, the endmembers and the variability with

$$\mathbf{A} \in \mathcal{A}_R = \big\{\mathbf{A} \in \mathbb{R}^{R \times N}\,\big|\, \mathbf{a}_n \in \mathcal{S}_R, \text{ for } n = 1, \dots, N\big\} \quad (8)$$

$$\mathbf{dM} \in \mathcal{D} = \mathcal{B}_{\mathrm{F}}(\mathbf{0}, \sigma) \cap \big\{\mathbf{dM} \mid \big\|\mathbb{E}\big[\mathbf{dM}\big]\big\|_{\mathrm{F}} \leq \kappa\big\}. \quad (9)$$

The parameters α, β and γ ensure a trade-off between the data fitting term and the penalties. In practice, g is approximated at time t by an upper bound $\hat{g}_t$ given by a stochastic approximation [11]

$$\hat{g}_t(\mathbf{M}) = \frac{1}{2t}\sum_{i=1}^{t}\|\mathbf{Y}_i - (\mathbf{M} + \mathbf{dM}_i)\mathbf{A}_i\|_{\mathrm{F}}^2 + \beta\Psi(\mathbf{M})$$

$$= \frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{2}\|\mathbf{M}\mathbf{A}_i\|_{\mathrm{F}}^2 - \langle \mathbf{Y}_i - \mathbf{dM}_i\mathbf{A}_i, \mathbf{M}\mathbf{A}_i\rangle\right) \\ + \beta\Psi(\mathbf{M}) + c \qquad (10)$$

$$= \frac{1}{t}\left[\frac{1}{2}\operatorname{Tr}(\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{C}_t) + \operatorname{Tr}(\mathbf{M}^{\mathrm{T}}\mathbf{D}_t)\right] + \beta\Psi(\mathbf{M}) + c$$

where $\langle \mathbf{X}, \mathbf{Y}\rangle = \operatorname{Tr}(\mathbf{X}^{\mathrm{T}}\mathbf{Y})$, c is a constant independent from $\mathbf{M}$ and

$$\mathbf{C}_t = \sum_{i=1}^{t}\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}}, \quad \mathbf{D}_t = \sum_{i=1}^{t}(\mathbf{dM}_i\mathbf{A}_i - \mathbf{Y}_i)\mathbf{A}_i^{\mathrm{T}}. \quad (11)$$

Besides, $\mathcal{D}$ is approximated by

$$\mathcal{D}_t = \mathcal{B}_{\mathrm{F}}(\mathbf{0}, \sigma) \cap \big\{\mathbf{dM} \mid \|\mathbf{dM} + \mathbf{E}_{t-1}\|_{\mathrm{F}} \leq t\kappa\big\} \quad (12)$$

with

$$\mathbf{E}_t = \sum_{i=1}^{t}\mathbf{dM}_i. \qquad (13)$$

Examples of penalizations that will be considered in this study are detailed in the following paragraphs.

Algorithm 1: Online unmixing algorithm.

Data: $\mathbf{M}^{(0)}, \mathbf{A}_0, \mathbf{dM}_0, \alpha > 0, \beta > 0, \gamma > 0, \xi \in\,]0,1]$
begin
  $\mathbf{C}_0 \leftarrow \mathbf{0}_{R,R}$;
  $\mathbf{D}_0 \leftarrow \mathbf{0}_{L,R}$;
  $\mathbf{E}_0 \leftarrow \mathbf{0}_{L,R}$;
  for $t = 1$ **to** T **do**
    a   Random selection of an image $\mathbf{Y}_t$ (random permutation of the image sequence);
    // Abundance and variability estimation by PALM [19], cf. §III-B1
    b $(\mathbf{A}_t, \mathbf{dM}_t) \in \underset{(\mathbf{A},\mathbf{dM}) \in \mathcal{A}_R \times \mathcal{D}_t}{\arg\min} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}, \mathbf{dM})$;
    $\mathbf{C}_t \leftarrow \xi\mathbf{C}_{t-1} + \mathbf{A}_t\mathbf{A}_t^{\mathrm{T}}$;
    $\mathbf{D}_t \leftarrow \xi\mathbf{D}_{t-1} + (\mathbf{dM}_t\mathbf{A}_t - \mathbf{Y}_t)\mathbf{A}_t^{\mathrm{T}}$;
    $\mathbf{E}_t \leftarrow \xi\mathbf{E}_{t-1} + \mathbf{dM}_t$;
    // Endmember update [11, Algo. 2], cf. §III-B2
    c $\mathbf{M}^{(t)} \leftarrow \underset{\mathbf{M} \in \mathcal{M}}{\arg\min}\, \hat{g}_t(\mathbf{M})$;

Result: $\mathbf{M}^{(T)}, (\mathbf{A}_t)_{t=1,\cdots,T}, (\mathbf{dM}_t)_{t=1,\cdots,T}$

---

1) Abundance penalization: In this work, the abundance penalization Φ has been chosen to promote temporally smooth abundances – in the ℓ_2-norm sense – between two consecutive images, leading to

$$\Phi(\mathbf{A}_t) = \frac{1}{2}\|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\mathrm{F}}^2. \qquad (14)$$

As long as Φ satisfies the regularity condition given in Paragraph III-C, any other type of prior knowledge relative to the abundances can be incorporated into the proposed method.

2) Endmember penalization: Classical endmember penalizations found in the literature consist in constraining the size of the $(R-1)$-simplex whose vertices are the endmembers. In this paper, we consider the mutual distance between each endmember introduced in [17], [18], defined as

$$\Psi(\mathbf{M}) = \frac{1}{2}\sum_{i=1}^{R}\left(\sum_{\substack{j=1 \\ j \neq i}}^{R}\|\mathbf{m}_i - \mathbf{m}_j\|_2^2\right) = \frac{1}{2}\sum_{r=1}^{R}\|\mathbf{M}\mathbf{G}_r\|_{\mathrm{F}}^2 \quad (15)$$

where

$$\mathbf{G}_r = -\mathbf{I}_R + \mathbf{e}_r\mathbf{1}_R^T \qquad (16)$$

and $\mathbf{e}_r$ denotes the rth canonical basis vector of $\mathbb{R}^R$.

3) Variability penalization: Assuming that the spectral variation between two consecutive images is *a priori* temporally smooth, we consider the following ℓ_2-norm penalization

$$\Upsilon(\mathbf{dM}_t) = \frac{1}{2}\|\mathbf{dM}_t - \mathbf{dM}_{t-1}\|_{\mathrm{F}}^2. \qquad (17)$$

Similarly, any other type of prior knowledge relative to the variability can be considered as long as Υ satisfies the regularity condition given in Paragraph III-C.

III. A TWO-STAGE STOCHASTIC PROGRAM

A. Two-stage stochastic program: general principle

The following lines briefly recall the main ideas presented in the introduction of [10]. A two-stage stochastic program is generally expressed as

$$\min_{\mathbf{M}} \mathbb{E}_{\mathbf{Y},\mathbf{Z}}\Big[f\big(\mathbf{Y}, \mathbf{M}, \mathbf{Z}\big)\Big] \text{ s.t. } \mathbf{M} \in \mathcal{M}, \text{ with } \mathbf{Z} \in \mathcal{Z}. \quad (18)$$

At the first stage, $\mathbf{M}$ must be chosen before any new data $\mathbf{Y}$ is available. At the second-stage, when $\mathbf{M}$ has been fixed

---

Algorithm 2: Abundance and variability estimation using PALM.

Data: $\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(0)}, \mathbf{dM}_t^{(0)}, \mathbf{E}_{t-1}$
begin
  $k \leftarrow 0$;
  while *stopping criterion not satisfied* **do**
    // Abundance update
    $\mathbf{A}_t^{(k+1)} \leftarrow \text{Update}\left(\mathbf{A}_t^{(k)}\right)$;   // cf. (25)
    // Variability update
    $\mathbf{dM}_t^{(k+1)} \leftarrow \text{Update}\left(\mathbf{dM}_t^{(k)}\right)$;   // cf. (28)
    $k \leftarrow k + 1$;
  $\mathbf{A}_t \leftarrow \mathbf{A}_t^{(k)}, \mathbf{dM}_t \leftarrow \mathbf{dM}_t^{(k)}$;
Result: $(\mathbf{A}_t, \mathbf{dM}_t)$

---

and a new data is acquired, the second-stage variable $\mathbf{Z}$ is computed as the solution (if it is unique and well defined) to the optimization problem

$$\min_{\mathbf{Z} \in \mathcal{Z}} f(\mathbf{Y}, \mathbf{M}, \mathbf{Z}). \tag{19}$$

Given an independent and identically distributed (i.i.d) T-sample $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$, problem (18) can be approximated by the sample average approximation (SAA)

$$\min_{\mathbf{M}, \mathbf{Z}_1, \ldots, \mathbf{Z}_T} \frac{1}{T} \sum_{t=1}^T f(\mathbf{Y}_t, \mathbf{M}, \mathbf{Z}_t), \text{ s.t. } \mathbf{M} \in \mathcal{M}, \mathbf{Z}_t \in \mathcal{Z}. \tag{20}$$

Moreover, when the second-stage (19) admits a unique solution, (20) can be rewritten as

$$\min_{\mathbf{M} \in \mathcal{M}} \frac{1}{T} \sum_{t=1}^T h(\mathbf{Y}_t, \mathbf{M}) \tag{21}$$

$$h(\mathbf{Y}_t, \mathbf{M}) = \min_{\mathbf{Z} \in \mathcal{Z}} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{Z}) \tag{22}$$

which is the SAA corresponding to

$$\min_{\mathbf{M} \in \mathcal{M}} \mathbb{E}_{\mathbf{Y}}\left[h(\mathbf{Y}, \mathbf{M})\right] \tag{23}$$

$$h(\mathbf{Y}, \mathbf{M}) = \min_{\mathbf{Z} \in \mathcal{Z}} f(\mathbf{Y}, \mathbf{M}, \mathbf{Z}) \tag{24}$$

where the two stages explicitly appear. However, f defined in (7) is non-convex with respect to $\mathbf{Z} = (\mathbf{A}, \mathbf{dM})$, where $\mathcal{Z} = \mathcal{A}_R \times \mathcal{D}$. Thus, problem (19) does not admit a unique global minimum, and existing algorithms will at most provide a critical point of $f(\mathbf{Y}, \mathbf{M}, \cdot) + \iota_{\mathcal{Z}}$, where $\iota_{\mathcal{Z}}$ denotes the indicator function of the set $\mathcal{Z}$. In this specific case, a new convergence framework based on a generalized equation has been developed in [10]. Such a framework enables a convergence result in terms of a critical point $\{\mathbf{M}, \mathbf{Z}_1, \ldots, \mathbf{Z}_T\}$ of (20) to be obtained. However, the significant size of the SAA problem (20) in our case is generally too expensive from a computational point of view. To alleviate this problem, we propose to slightly adapt the work developed in [11] to propose an online estimation algorithm described in Algo. 1. This algorithm has the same convergence property as [11] provided the non-convex function $f(\mathbf{Y}, \mathbf{M}, \cdot) + \iota_{\mathcal{Z}}$ exclusively admits locally unique critical points. Further details are given in Paragraph III-C.

B. Parameter estimation

Whenever an image $\mathbf{Y}_t$ has been received, the abundances and variability are estimated by a proximal alternating linearized minimization (PALM) algorithm [19], which is

---

Algorithm 3: Endmember estimation.

Data: $\mathbf{M}^{(t,0)} = \mathbf{M}^{(t-1)}, \mathbf{C}_t, \mathbf{D}_t$
begin
  $k \leftarrow 0$;
  while *stopping criterion not satisfied* **do**
    // Endmember update
    $\mathbf{M}^{(t,k+1)} \leftarrow \text{Update}\left(\mathbf{M}^{(t,k)}\right)$;   // cf. (31)
    $k \leftarrow k + 1$;
  $\mathbf{M}^{(t)} \leftarrow \mathbf{M}^{(t,k)}$;
Result: $\mathbf{M}^{(t)}$

---

guaranteed to converge to a critical point $(\mathbf{A}^*, \mathbf{dM}^*)$ of $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot) + \iota_{\mathcal{A}_R \times \mathcal{D}_t}$. The endmembers are then updated by proximal gradient descent steps, similarly to [11]. Further details on the projections involved in this section are given in Appendix A.

1) Abundance and variability estimation: A direct application of [19] under the constraints (3) leads to the following abundance update rule

$$\mathbf{A}_t^{(k+1)} = \mathcal{P}_{\mathcal{A}_R}\left(\mathbf{A}_t^{(k)} - \frac{1}{L_{1t}^{(k)}} \nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(k)}, \mathbf{dM}_t^{(k)})\right) \tag{25}$$

where $L_{1t}^{(k)}$ is the Lipschitz constant of $\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \cdot, \mathbf{dM}_t^{(k)})$ and

$$\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \alpha(\mathbf{A}_t - \mathbf{A}_{t-1}) + (\mathbf{M}^{(t)} + \mathbf{dM}_t)^{\mathrm{T}}\left[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t\right] \tag{26}$$

$$L_{1t}^{(k)} = \left\|(\mathbf{M}^{(t)} + \mathbf{dM}_t^{(k)})^{\mathrm{T}}(\mathbf{M}^{(t)} + \mathbf{dM}_t^{(k)}) + \alpha \mathbf{I}_R\right\|_{\mathrm{F}}. \tag{27}$$

Note that the projection $\mathcal{P}_{\mathcal{A}_R}$ can be exactly computed using the algorithms proposed in [20], [21]. Similarly, the update rule for the variability terms is

$$\mathbf{dM}_t^{(k+1)} = \mathcal{P}_{\mathcal{D}_t}\left(\mathbf{dM}_t^{(k)} - \frac{1}{L_{2t}^{(k)}} \nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(k+1)}, \mathbf{dM}_t^{(k)})\right) \tag{28}$$

where $L_{2t}^{(k)}$ is the Lipschitz constant of $\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(k+1)}, \cdot)$ and

$$\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \gamma(\mathbf{dM}_t - \mathbf{dM}_{t-1}) + \left[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t\right]\mathbf{A}_t^{\mathrm{T}} \tag{29}$$

$$L_{2t}^{(k)} = \left\|\mathbf{A}_t^{(k+1)}\mathbf{A}_t^{(k+1)\mathrm{T}} + \gamma \mathbf{I}_R\right\|_{\mathrm{F}}. \tag{30}$$

Note that the projection $\mathcal{P}_{\mathcal{D}_t}$ can be efficiently approximated using the Dykstra algorithm (see [22]–[24]). The resulting algorithm is summarized in Algo. 2.

2) Endmember estimation: Similarly to III-B1, a direct application of the method detailed in [11], [19] yields

$$\mathbf{M}^{(t,k+1)} = \mathcal{P}_+\left(\mathbf{M}^{(t,k)} - \frac{1}{L_{3t}} \nabla_{\mathbf{M}} \hat{g}_t(\mathbf{M}^{(t,k)})\right) \tag{31}$$

where $\mathcal{P}_+$ is the projector on $\{\mathbf{X} | \mathbf{X} \succeq \mathbf{0}_{L,R}\}$ and L_{3t} denotes the Lipschitz constant of $\nabla_{\mathbf{M}} \hat{g}_t(\mathbf{M}^{(t,k)})$. Note that

$$\nabla_{\mathbf{M}}\hat{g}_t(\mathbf{M}) = \mathbf{M}\left(\frac{1}{t}\mathbf{C}_t + \beta\sum_{r=1}^{R}\mathbf{G}_r\mathbf{G}_r^{\mathrm{T}}\right) - \frac{1}{t}\mathbf{D}_t \qquad (32)$$

$$L_{3t} = \left\|\frac{1}{t}\mathbf{C}_t + \beta\sum_r\mathbf{G}_r\mathbf{G}_r^{\mathrm{T}}\right\|_{\mathrm{F}}. \qquad (33)$$

The resulting algorithm is summarized in Algo. 3.

C. Convergence guarantee

To ensure the convergence of the generated endmember sequence $(\mathbf{M}^{(t)})_t$ towards a critical point of the problem (18), we make the following assumptions.

Assumption 1. The quadratic functions $\hat{g}_t$ are strictly convex and admit a Hessian matrix lower-bounded in norm by a constant $\mu_{\mathbf{M}} > 0$.

Assumption 2. The penalty functions Φ, Ψ and Υ are gradient Lipschitz continuous with Lipschitz constant c_Φ, c_Ψ and c_Υ respectively. In addition, Φ and Υ are assumed to be twice continuously differentiable.

Assumption 3. The function $f(\mathbf{Y}_t, \cdot, \cdot, \cdot)$ is twice continuously differentiable. The Hessian matrix of $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot)$ – denoted by $\mathbf{H}_{(\mathbf{A}, \mathbf{dM})} f$ – is invertible at each critical point $(\mathbf{A}_t^*, \mathbf{dM}_t^*) \in \mathcal{Q}(\mathbf{Y}_t, \mathbf{M})$.

In practice, Assumption 1 may be enforced by adding a penalization term $\frac{\mu_{\mathbf{M}}}{2}\|\mathbf{M}\|_{\mathrm{F}}^2$ to the objective function $\hat{g}_t$, where $\mu_{\mathbf{M}}$ is a small positive constant. Note that $\mu_{\mathbf{M}}$ is only a technical guarantee used in the convergence proof reported in Appendix C, which should not be computed explicitly to be able to run the algorithm. Assumption 2 is only included here for the sake of completeness, in case other penalizations than those given in Section II are considered. Indeed, this assumption is obviously satisfied by the penalizations mentioned in this work. Assumption 3, crucial to Proposition 1, is further discussed in Appendix B to ease the reading of this paper.

By adapting the arguments used in [11], the convergence property summarized in Proposition 1 can be obtained.

Proposition 1 (Convergence of $(\mathbf{M}^{(t)})_t$, [11])**.** *Under the assumptions 1,2 and 3, the distance between $\mathbf{M}^{(t)}$ and the set of critical points of the hyperspectral unmixing problem (6) converges almost surely to 0 when t tends to infinity.*

Proof. See Appendix C. □

D. Computational complexity

Dominated by matrix-product operations, the per image overall complexity of the proposed method is of the order

$$\mathcal{O}\Big\{\big[LR(N + N_{\mathrm{iter}}^D) + R^2(L + N)\big]N_{\mathrm{iter}}^P + N_{\mathrm{iter}}LR^2\Big\}$$

where N_{iter}^D, N_{iter}^P, N_{iter} denote the number of iterations for the Dykstra algorithm involved in the variability projection (28), the PALM algorithm and the endmember update respectively. To be more explicit, the computation time for one image of size 100×100 composed of $L = 173$ bands is approximately 6 s for a MATLAB implementation with an Intel(R) Core(TM) i5-4670 CPU @ 3.40GHz. Note that the PALM iterations
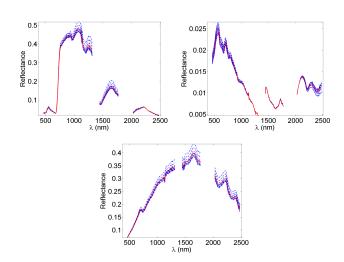

Fig. 1. Reference endmembers (red lines) and the corresponding instances under spectral variability (blue lines) involved in the synthetic HS images.

TABLE II
PARAMETERS USED IN THE EXPERIMENTS.

	Synthetic data	Real data
σ^2	1	1
κ^2	0.1	0.01
α	10^{-4}	0
β	10^{-3}	10^{-4}
γ	3×10^{-5}	0
N_{iter}^D	50	50
N_{iter}^P	50	50
N_{iter}	50	50
N_{epochs}	10	10
ξ	0.98	0.98

(Algo. 2) and the endmember updates (Algo. 3) can be parallelized if needed due to the separability of the objective function f chosen (separability with respect to the column of the abundance matrix, and with respect to the rows of the endmember and variability matrices).

IV. EXPERIMENT WITH SYNTHETIC DATA

This section considers an HS image sequence composed of 10 images of size 98×102, each image composed of 173 bands. The images correspond to linear mixtures of 3, 6 and 10 endmembers affected by smooth time-varying variability. The synthetic abundance maps of this scenario vary smoothly from one image to another. Note that the pure pixel assumption is not satisfied for all images of the experiment with $R = 3$ endmembers in order to assess the algorithm performance in a challenging scenario. The synthetic linear mixtures have been corrupted by additive white Gaussian noise to ensure a resulting signal-to-noise ratio of SNR $= 30$ dB. Additional results for mixtures corrupted by colored Gaussian noise are available in [25, App. D].

In order to introduce controlled spectral variability, the endmembers involved in the mixtures have been generated using the product of reference endmembers with randomly generated piecewise-affine functions as in [7]. The corresponding perturbed endmembers used in the experiment are
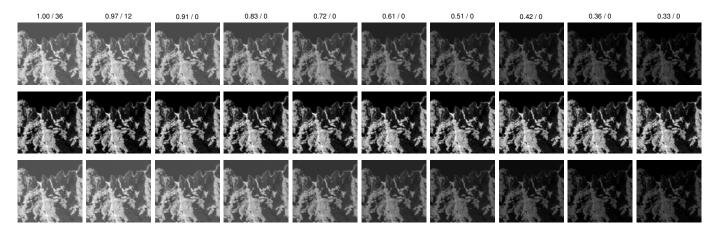
Fig. 2. Abundance maps of the first endmember used in the synthetic mixtures (theoretical abundances on the first line, VCA/FCLS on the second line, proposed method on the third line). The top line indicates the theoretical maximum abundance value and the true number of pixels whose abundance is greater than 0.95 for each time instant.
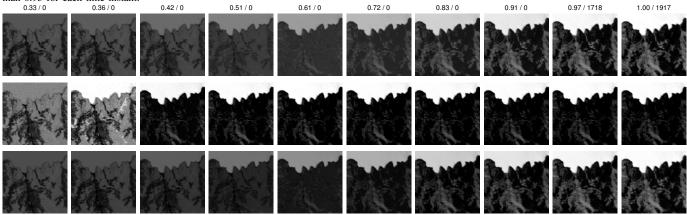

Fig. 3. Abundance maps of the second endmember used in the synthetic mixtures (theoretical abundances on the first line, VCA/FCLS on the second line, proposed method on the third line). The top line indicates the theoretical maximum abundance value and the true number of pixels whose abundance is greater than 0.95 for each time instant.

depicted in Fig. 1. Note that different affine functions have been considered at each time instant for each endmember.

A. Compared methods

The results of the proposed algorithm have been compared to those obtained with several classical linear unmixing methods performed individually on each image of the time series. The methods are recalled below with their most relevant implementation details. All the methods requiring an appropriate initialization have been initialized with VCA/FCLS.

1) VCA/FCLS (no variability): for each image, the endmembers are first extracted using the vertex component analysis (VCA) [12] which requires pure pixels to be present in the analyzed images. The abundances are then estimated for each pixel by solving a Fully Constrained Problem (FCLS) with ADMM [14];

2) SISAL/FCLS (no variability): the endmembers are first extracted using the simplex identification via split augmented Lagrangian (SISAL) [15]. Note that the pure pixel assumption is not required to apply this method. The tolerance for the stopping rule has been set to 10^{-3}. The abundances are then estimated for each pixel by FCLS;

3) $\ell_{1/2}$ NMF (no variability): the algorithm described in [16] is applied to each image, with a stopping criterion

set to 10^{-3} and a maximum of 300 iterations. The regularization parameter has been set as in [16];

4) BCD/ADMM: the algorithm described in [7] is applied to each image with a stopping criterion set to 10^{-3}. The endmember regularization recalled in (15) has been used, with a parameter set to the same value as the one used for the proposed method. The abundance regularization parameter (spatial smoothness) has been set to 10^{-4}, and the variability regularization parameter has been set to 1;

5) Proposed method: endmembers are initialized with VCA applied to the union of the pixels belonging to the $R-1$ convex hull of each image. The abundances are initialized by FCLS, and the variability matrices are initialized with all their entries equal to 0. Whenever the algorithm is applied to a previously processed image, the previous abundance and variability estimates are taken as a warm-restart. Algo. 2 (PALM algorithm) is stopped after N_{iter}^P iterations and the Dykstra algorithm used to compute the projection in (28) is iterated N_{iter}^D times. Moreover, Algo. 3 is stopped after N_{iter} iterations. Finally, Algo. 1 is stopped after N_{epochs} cycles – referred to as epochs – on the randomly permuted training set to approximately obtain i.i.d. samples [11]. In particular, the number of cycles N_{epochs} and sub-iterations N_{iter} have been empirically
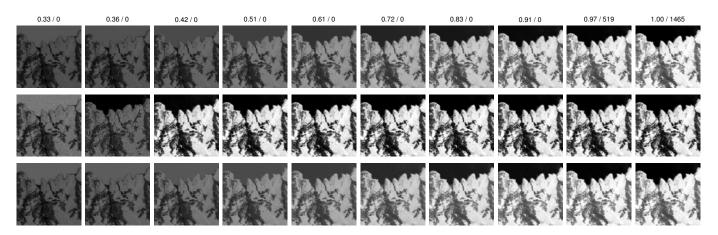
Fig. 4. Abundance maps of the third endmember used in the synthetic mixtures (theoretical abundances on the first line, VCA/FCLS on the second line, proposed method on the third line). The top line indicates the theoretical maximum abundance value and the true number of pixels whose abundance is greater than 0.95 for each time instant.
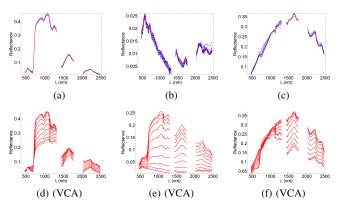

Fig. 5. Estimated endmembers on the synthetic hyperspectral time series (PLMM endmembers in red with variability in blue dotted lines on the first line, VCA-extracted endmembers on the second line, SISAL endmembers are omitted since very similar to those obtained with VCA).

chosen to obtain a compromise between the estimation accuracy and the implied computational cost. We also included a constant forgetting factor $\xi \in (0, 1)$ in order to slowly forget the past data. The closer to one ξ is, the more slowly the past data are forgotten.

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

$$\text{aSAM}(\mathbf{M}) = \frac{1}{R} \sum_{r=1}^{R} \arccos \left(\frac{\mathbf{m}_r^{\text{T}} \widehat{\mathbf{m}}_r}{\|\mathbf{m}_r\|_2 \|\widehat{\mathbf{m}}_r\|_2} \right) \quad (34)$$

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

$$\text{GMSE}(\mathbf{A}) = \frac{1}{TRN} \sum_{t=1}^{T} \|\mathbf{A}_t - \widehat{\mathbf{A}}_t\|_{\text{F}}^2 \quad (35)$$

$$\text{GMSE}(\mathbf{dM}) = \frac{1}{TLR} \sum_{t=1}^{T} \|\mathbf{dM}_t - \widehat{\mathbf{dM}}_t\|_{\text{F}}^2. \quad (36)$$

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

$$\text{RE} = \frac{1}{TLN} \sum_{t=1}^{T} \|\mathbf{Y}_t - \widehat{\mathbf{Y}}_t\|_{\text{F}}^2 \quad (37)$$

where $\widehat{\mathbf{Y}}_t$ is the matrix formed of the pixels reconstructed with the parameters estimated for the image t.

B. Results

The parameters used for the proposed algorithm, which have been adjusted by cross-validation, are detailed in Table II. For the dataset associated with mixtures of $R = 3$ endmembers, the abundance maps obtained by the proposed method are compared to those of VCA/FCLS in Figs. 2 to 4, whereas the corresponding endmembers are displayed in Fig. 5. The abundance maps obtained by SISAL/FCLS, $\ell_{1/2}$ NMF and BCD/ADMM, somewhat similar to those obtained by VCA/FCLS, are included in a separate technical report [25], along with a more detailed version of Table III and the endmembers extracted by all the unmixing strategies. The performance of the unmixing methods is finally reported in Table III, leading to the following conclusions.

- The proposed method is more robust to the absence of pure pixels in some images than both VCA/FCLS and SISAL/FCLS. Note that $\ell_{1/2}$ NMF and BCD/ADMM converge to poor local optima, which directly results from the poor performance of VCA in this specific context. On the contrary, the estimated abundances obtained with the proposed method (second line of Figs. 2 to 4) are closer to the ground truth (first line) than VCA/FCLS (third line). This observation is confirmed by the results given in Table III;

- The proposed method provides competitive unmixing results while allowing temporal endmember variability to be estimated for each endmember (see Fig. 5);

- The abundance GMSEs and the REs estimated with the proposed method are lower or comparable to those obtained with VCA/FCLS and SISAL/FCLS applied to each image individually (see Table III), without introducing much more degrees of freedom into the underlying model when compared to BCD/ADMM;

- Even though the performance of the proposed method degrades with the number of endmembers, the results remain better or comparable to those of the other methods.

Whenever an endmember is scarcely present in one of the images, the proposed method outperforms VCA/FCLS as can be seen in Figs. 2 to 4. Note that the maximum theoretical abundance value and the number of pixels whose abundances are greater than 0.95 are mentioned on the top line of Figs.

TABLE III
SIMULATION RESULTS ON SYNTHETIC DATA (aSAM(M) IN (°),
GMSE(A)$\times 10^{-2}$, GMSE(dM)$\times 10^{-4}$, RE $\times 10^{-4}$, TIME IN (S)).

		aSAM(M)	GMSE(A)	GMSE(dM)	RE	time
$R = 3$	VCA/FCLS	15.76	4.22	/	0.413	**1**
	SISAL/FCLS	15.88	3.68	/	**0.375**	5
	$\ell_{1/2}$ NMF	18.54	8.20	/	0.123	226
	BCD/ADMM	15.91	3.47	4.03	0.282	379
	Proposed	**1.88**	**0.23**	1.02	**0.375**	168
$R = 6$	VCA/FCLS	2.14	**0.14**	/	1.48	**4**
	SISAL/FCLS	1.67	0.83	/	**1.20**	5
	$\ell_{1/2}$ NMF	3.41	0.45	/	1.53	332
	BCD/ADMM	2.27	0.29	1.31	1.20	1066
	Proposed	**1.49**	0.17	2.69	1.22	344
$R = 10$	VCA/FCLS	3.52	7.24	/	4.63	**5**
	SISAL/FCLS	9.53	3.32	/	**1.67**	6
	$\ell_{1/2}$ NMF	5.58	7.03	/	3.90	279
	BCD/ADMM	3.27	6.45	7.2	1.70	735
	Proposed	**2.83**	**0.43**	8.9	1.99	204

2 to 4, to assess the difficulty of recovering each endmember in each image. This result was expected, since VCA is a pure pixel-based unmixing method.

C. Hyper-parameter influence on the reconstruction error

Considering the significant number of hyper-parameters to be tuned (i.e., $\alpha, \beta, \gamma, \sigma, \kappa$), a full sensitivity analysis is a challenging task, which is further complexified by the non-convex nature of the problem considered. To alleviate this issue, each parameter has been individually adjusted while the others were set to a priori reasonable values (i.e., $(\alpha, \beta, \gamma, \sigma^2, \kappa^2) = (10^{-2}, 10^{-4}, 10^{-4}, \hat{\sigma}^2, 10^{-3})$, where $\hat{\sigma}^2 = 0.0372$ denotes the theoretical average energy of the variability introduced in the synthetic dataset used for this analysis). The appropriateness of a given range of values has been evaluated in terms of the RE of the recovered solution. The results reported in Fig. 6 suggest that the proposed method is relatively robust to the choice of the hyper-parameters. More precisely, as can be seen in Figs. 6b and 6c, only β and γ may induce oscillations (of very small amplitude) in the RE. Based on this analysis, it is interesting to note that the interval $[2 \times 10^{-3}, 10^{-2}]$ can be chosen in practice to obtain reasonable reconstruction errors.

To conclude, the two following remarks can be made on the choice of σ and κ:

- the value chosen for σ results from an empirical compromise between the risk to capture noise into the variability terms (σ too large) and the risk to lose information (σ too small). The sensitivity analysis conducted in Fig. 6d shows that $\sigma^2 \in [10^{-1}, 1]$ provides interesting results for this experiment;
- κ should be set to a value ensuring that $\mathbf{M}$ reflects the average spectral behavior of the perturbed endmembers. Fig. 6e shows that $\kappa^2 \in [10^{-3}, 1]$ provides interesting results for the synthetic dataset used in the experiment.

V. EXPERIMENT WITH REAL DATA

A. Description of the dataset

The proposed algorithm has been applied to real HS images acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the Lake Tahoe region (California, United
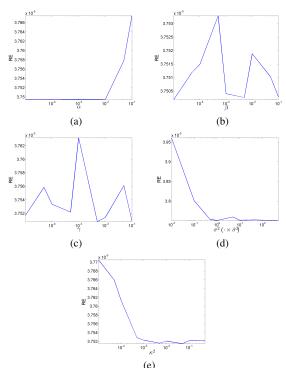

Fig. 6. Sensitivity analysis of the reconstruction error RE with respect to the tuning of the algorithm hyper-parameters ($\hat{\sigma}^2 = 0.0372$ denotes the theoretical average energy of the variability introduced in the synthetic dataset used for this analysis).

States of America) between 2014 and 2015[3]. Water absorption bands were removed from the 224 spectral bands, leading to 173 exploitable bands. In absence of any ground truth, the sub-scene of interest (150×110), partly composed of a lake and a nearby field, has been unmixed with $R = 3$, 4 and 5 endmembers to obtain a compromise between the results of HySime [26], those of the recently proposed eigen-gap approach (EGA) [27] (see Table IV), and the consistency of the resulting abundance maps. The parameters used for the proposed approach are given in Table II, and the other methods have been run with the same parameters as in Section V. Note that a 4×4 patch composed of outliers has been manually removed from the last image of the sequence prior to the unmixing procedure.

B. Results

Since no ground truth is available, the algorithm performance is evaluated in terms of the reconstruction error defined in (37). Only the more consistent abundance maps and endmembers obtained for $R = 3$ are presented in Figs. 8 to 11 due to space constraints. Complementary results are available in [25]. The proposed method provides comparable reconstruction errors (see Table V), yields more consistent abundance maps when compared to VCA/FCLS and SISAL/FCLS especially for the soil and the vegetation for a somewhat reasonable computational cost. In particular, note that the estimated vegetation abundance map of the fourth image depicted in Fig. 10 (area delineated in red) presents

---

[3]The images used in this experiment are freely available from the online AVIRIS flight locator tool at http://aviris.jpl.nasa.gov/alt_locator/.

(a) 04/10/2014 (b) 06/02/2014 (c) 09/19/2014 (d) 11/17/2014 (e) 04/29/2015

Fig. 7. Scenes used in the experiment, given with their respective acquisition date.
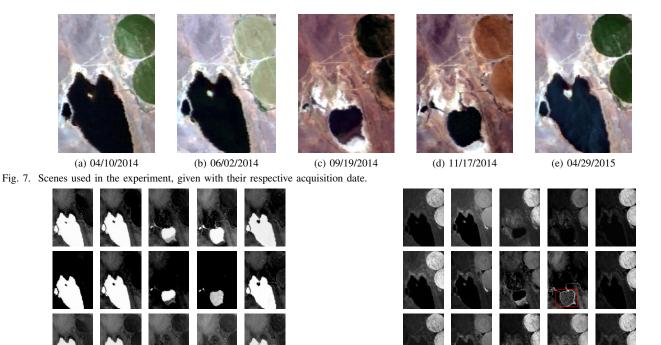

Fig. 8. Water abundance maps (proposed method on the first line, VCA/FCLS on the second line, SISAL/FCLS on the third line).
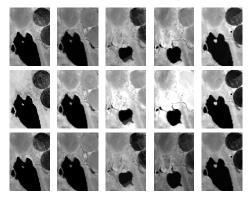

Fig. 9. Soil abundance maps (proposed method on the first line, VCA/FCLS on the second line, SISAL/FCLS on the third line).

Fig. 10. Vegetation abundance maps (proposed method on the first line, VCA/FCLS on the second line, SISAL/FCLS on the third line). The region delineated in red, where almost no vegetation is supposed to be present, reveals that the water endmember extracted by VCA has been split into two parts. This observation is further confirmed in Figs. 11d and 11f.

TABLE IV
ENDMEMBER NUMBER R ESTIMATED ON EACH IMAGE OF THE REAL
DATASET BY HYSIME [26] AND EGA [27].

	04/10/2014	06/02/2014	09/19/2014	11/17/2014	04/29/2015
HySime [26]	16	21	19	21	22
EGA [27]	3	5	4	3	3

significant errors when visually compared to the corresponding RGB image in Fig. 7d. These errors can be explained by the fact that the water endmember extracted by VCA has been split into two parts as can be seen in Figs. 11d and 11f (see signatures given in black). Indeed, the VCA algorithm cannot detect the scarcely present vegetation. On the contrary, the joint exploitation of multiple images enables the faint traces of dry vegetation to be captured. Albeit impacted by the results of VCA/FCLS (used as initialization), the performance of $\ell_{1/2}$ NMF and BCD/ADMM remains satisfactory on each image of the sequence since they tend to correct the endmember errors induced by VCA. However, $\ell_{1/2}$ NMF produces undesirable endmembers with an amplitude significantly greater than 1 on the 4th image (Fig. 7d). Besides, BCD/ADMM yields very low reconstruction errors at the price of a computational cost which may become prohibitive for extended image sequences. The figures related to $\ell_{1/2}$ NMF and BCD/ADMM are available

in the associated report [25] due to space constraints.

Furthermore the instantaneous variability energy (computed as $\|\mathbf{dm}_{rt}\|_2^2/L$ for $r = 1, \ldots, R$ and $t = 1 \ldots, T$) can reveal which endmember deviates the most from its average spectral behavior. In this experiment, the soil and the vegetation signatures – which seem to vary the most over time (see Fig. 7) – are found by the proposed method to be affected by the most significant variability level (see Table VI). In this experiment, a significant increase can be observed in the endmember variability energy over the last three images of the sequence (see Table VI), suggesting that the endmembers are apparently better represented in the two first images of the sequence (see Fig. 7). This observation suggests the proposed method captures the average endmember spectral behavior and enables the time at which the greatest spectral changes occur to be identified. However, a detailed analysis of this observation is out of the scope of the present paper.

(a) Water

(b) Soil

(c) Vegetation

(d) Water (VCA)

(e) Soil (VCA)

(f) Vegetation (VCA)

(g) Water (SISAL)

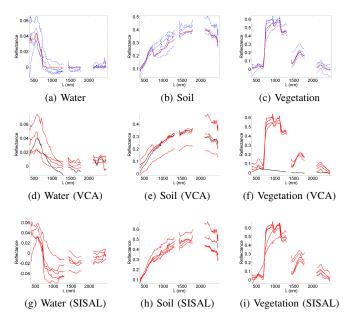(h) Soil (SISAL)

(i) Vegetation (SISAL)

Fig. 11. Endmembers and variability (endmembers in red lines, variability in blue dashed lines) recovered by the proposed method on the first line, VCA-extracted endmembers on the second line, SISAL-extracted endmembers on the third line). The endmembers given in black on the second line correspond to the endmembers identified by VCA on the image 7d, where the water endmember has been split into two parts (see Figs. 11d and 11f).

TABLE V
SIMULATION RESULTS ON REAL DATA (RE $\times 10^{-4}$).

		RE	time (s)
$R = 3$	VCA/FCLS	12.7	**2**
	SISAL/FCLS	0.87	3
	$\ell_{1/2}$ NMF	3.83	156
	BCD/ADMM	**0.37**	2449
	Proposed	1.04	134
$R = 4$	VCA/FCLS	43.8	**2**
	SISAL/FCLS	0.35	3
	$\ell_{1/2}$ NMF	16.0	163
	BCD/ADMM	**0.27**	4396
	Proposed	0.76	126
$R = 5$	VCA/FCLS	63.9	**2**
	SISAL/FCLS	0.17	4
	$\ell_{1/2}$ NMF	14.6	174
	BCD/ADMM	**0.098**	12511
	Proposed	0.17	128

VI. CONCLUSION AND FUTURE WORK

This paper introduced an online hyperspectral unmixing procedure accounting for endmember temporal variability based on the perturbed linear model considered in [7]. This algorithm was designed to unmix multiple HS images of moderate size, potentially affected by smoothly varying endmember perturbations. Indeed, the number of spurious local optima of the cost function used in this paper can significantly increase with the size of the images and the number of endmembers considered, which is a problem common to many blind source separation problems (such as the unmixing problem addressed in this paper). The underlying unmixing problem was formulated as a two-stage stochastic program solved by a stochastic approximation algorithm. Simulations conducted on synthetic and real data enabled the interest of the proposed

TABLE VI
EXPERIMENT WITH REAL DATA FOR $R = 3$: ENERGY OF THE VARIABILITY CAPTURED FOR EACH ENDMEMBER AT EACH TIME INSTANT ($\|\mathbf{dm}_{kt}\|_2^2 / L \times 10^{-5}$ FOR $k = 1, \ldots, R$, $t = 1, \ldots, T$).

	Water	Vegetation	Soil
04/10/2014	1.22	9.68	**11.51**
06/02/2014	1.44	11.85	**38.37**
09/19/2014	7.29	**11.41**	9.30
11/17/2014	2.77	**21.73**	16.55
04/29/2015	0.58	**106.03**	26.19

approach to be appreciated. Indeed, the proposed method compared favorably with established approaches performed independently on each image of the sequence while providing a relevant variability estimation. Assessing the robustness of the proposed technique with respect to estimation errors on the endmember number R and applying the proposed method to real dataset composed of a larger number of endmembers are interesting prospects for future work. Possible perspectives also include the extension of the method to account for spatial variability and applications to change detection problems. A distributed unmixing procedure is also under investigation to solve the resulting high dimensional problem.

APPENDIX A
PROJECTIONS INVOLVED IN THE PARAMETER UPDATES

The projections involved in the PALM algorithm [19] described in Algo. 2 are properly defined, since the associated constraint spaces are closed convex sets. More precisely,

- $\mathcal{D}_t$ is closed and convex as the (non-empty) intersection of two closed balls. The projection onto $\mathcal{D}_t$ can be approximated by the Dykstra algorithm [22], [24]. Besides, the projection on a Frobenius ball is given by [28]

$$\mathcal{P}_{\mathcal{B}_{\mathrm{F}}(\mathbf{X}, r)}(\mathbf{Y}) = \mathbf{X} + \min\left(1, \frac{r}{\|\mathbf{Y} - \mathbf{X}\|_{\mathrm{F}}}\right)(\mathbf{Y} - \mathbf{X}); \tag{38}$$

- projecting $\mathbf{M}$ onto $\mathbb{R}_+^{L \times R}$ is explicitly given by
$$\mathcal{P}_+(\mathbf{M}) = \max(\mathbf{0}_{L,R}, \mathbf{M}) \tag{39}$$

where the $\max$ is taken term-wise.

APPENDIX B
DISCUSSION ON ASSUMPTION 3

The Hessian matrix of $f(\mathbf{Y}, \mathbf{M}, \cdot, \cdot)$, denoted by $\mathbf{H}_{(\mathbf{A}, \mathbf{dM})} f$, is given by

$$\mathbf{H}_{(\mathbf{A}, \mathbf{dM})} f = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \end{bmatrix} \tag{40}$$

$$\widetilde{\mathbf{M}} = (\mathbf{M} + \mathbf{dM}) \tag{41}$$

$$\mathbf{H}_1 = \mathbf{I}_N \otimes (\widetilde{\mathbf{M}}^{\mathrm{T}} \widetilde{\mathbf{M}}), \quad \mathbf{H}_4 = (\mathbf{A}\mathbf{A}^{\mathrm{T}}) \otimes \mathbf{I}_L \tag{42}$$

$$\mathbf{H}_3 = \mathbf{H}_2^{\mathrm{T}} = \left\{ \mathbf{I}_R \otimes \left[-\mathbf{Y} + \widetilde{\mathbf{M}}\mathbf{A} \right] \right\} \mathbf{S}_{R,N} + [\mathbf{A} \otimes \widetilde{\mathbf{M}}]. \tag{43}$$

where $\mathbf{S}_{R,L}$ is the perfect shuffle matrix. The block matrix $\mathbf{H}_{(\mathbf{A}, \mathbf{dM})} f$ is invertible if, for instance, $\mathbf{H}_1$ and its Schur complement $\mathbf{S} = \mathbf{H}_4 - \mathbf{H}_3 \mathbf{H}_1^{-1} \mathbf{H}_2$ are invertible. In practice, $\mathbf{H}_1$ is generally invertible since $\mathbf{M} + \mathbf{dM}$ is full column rank. The invertibility of the Schur complement $\mathbf{S}$ can be ensured via an appropriate regularization term $\frac{\mu}{2} \|\mathbf{A}\|_{\mathrm{F}}^2$ added to the original objective f. Indeed, we first note that such a

perturbation regularizes the Hessian by modifying its diagonal block $\mathbf{H}_4$, replaced by $\mathbf{H}_4 + \mu\mathbf{I}$.

Denote by $\lambda_1 > \lambda_2 > \ldots > \lambda_r$ the ordered eigenvalues of $\mathbf{S}$, where r denotes the number of distinct eigenvalues. By the spectral theorem, there exists an orthogonal matrix (with respect to the canonical euclidean inner product) $\mathbf{Q}$ such that $\mathbf{S} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}\mathbf{Q}$, where $\mathbf{D}$ is a diagonal matrix composed of the λ_k. Note that each eigenvalue may have a multiplicity order greater than 1 with the adopted notations. If there exits k such that $\lambda_k = 0$, then $\lambda_{k+1} < 0$. Adding $\frac{\mu}{2}\|\mathbf{A}\|_{\mathrm{F}}^2$ to the original objective function, with $\mu < |\lambda_{k+1}|$, is then sufficient to ensure the invertibility of the Schur complement
$$(\mathbf{H}_4 - \mathbf{H}_3\mathbf{H}_1^{-1}\mathbf{H}_2) + \mu\mathbf{I} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}\mathbf{Q} + \mu\mathbf{I} = \mathbf{Q}^{\mathrm{T}}(\mathbf{D} + \mu\mathbf{I})\mathbf{Q}$$
associated to the new Hessian matrix, thus ensuring its invertibility.

APPENDIX C
CONVERGENCE PROOF

Largely adapted from [11], the following sketch of proof reduces to an adaptation of [11, Lemma 1, Proposition 1]. From this point, our problem exactly satisfies the assumptions required to apply the same arguments as in [11, Proposition 2, Proposition 3], leading to the announced convergence result.

Lemma 1 (Asymptotic variations of $\mathbf{M}_t$ [11]). *Under Assumptions 1 and 2, we have*
$$\left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_F = O\left(\frac{1}{t}\right) \text{ almost surely (a.s.).} \quad (44)$$
Proof. According to Assumption 1, $\hat{g}_t$ is strictly convex with a Hessian lower-bounded by a scalar $\mu_{\mathbf{M}} > 0$. Consequently, $\hat{g}_t$ satisfies the second-order growth condition
$$\hat{g}_t(\mathbf{M}^{(t+1)}) - \hat{g}_t(\mathbf{M}^{(t)}) \geq \mu_M \left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_{\mathrm{F}}^2. \quad (45)$$
Besides, since $\mathbf{M} \in [0,1]^{L \times R}$, we have $\|\mathbf{M}\|_{\mathrm{F}} \leq \sqrt{LR}$. Hence $\hat{g}_t$ is Lipschitz continuous with constant $c_t = \frac{1}{t}\left(\|\mathbf{D}_t\|_{\mathrm{F}} + \sqrt{LR}\|\mathbf{C}_t\|_{\mathrm{F}}\right) + \beta c_\Psi$. Indeed, given two matrices $\mathbf{M}_1, \mathbf{M}_2 \in [0,1]^{L \times R}$, we have
$$|\hat{g}_t(\mathbf{M}_1) - \hat{g}_t(\mathbf{M}_2)| \leq \beta|\Psi(\mathbf{M}_1) - \Psi(\mathbf{M}_2)| +$$
$$\frac{1}{t}\left|\frac{1}{2}\langle\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2, \mathbf{C}_t\rangle - \langle\mathbf{M}_1 - \mathbf{M}_2, \mathbf{D}_t\rangle\right|$$
$$\leq \beta c_\Psi \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}} + \frac{1}{t}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}\|\mathbf{D}_t\|_{\mathrm{F}} \quad (46)$$
$$+ \frac{1}{2t}\|\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2\|_{\mathrm{F}}\|\mathbf{C}_t\|_{\mathrm{F}}$$
where $\mathbf{C}_t$ and $\mathbf{D}_t$ were defined in (11). In addition
$$\|\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2\|_{\mathrm{F}} = \frac{1}{2}\|(\mathbf{M}_1 + \mathbf{M}_2)^{\mathrm{T}}(\mathbf{M}_1 - \mathbf{M}_2)$$
$$+ (\mathbf{M}_1 - \mathbf{M}_2)^{\mathrm{T}}(\mathbf{M}_1 + \mathbf{M}_2)\|_{\mathrm{F}} \quad (47)$$
$$\leq 2\sqrt{LR}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}$$
hence
$$|\hat{g}_t(\mathbf{M}_1) - \hat{g}_t(\mathbf{M}_2)| \leq c_t \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}. \quad (48)$$
Combining (45) and (48), we have
$$\left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_{\mathrm{F}} \leq \frac{c_t}{\mu_{\mathbf{M}}}. \quad (49)$$
Since the data, the abundances and the variability are respectively contained in compact sets, $\mathbf{C}_t$ and $\mathbf{D}_t$ are (almost surely) bounded, thus: $c_t = O\left(\frac{1}{t}\right)$ a.s. $\qquad\square$

Proposition 2 (Adapted from [11]). *We assume that the requirements in Assumption 1 to 3 are satisfied. Let $(\mathbf{Y}_t, \mathbf{M})$ be an element of $\mathcal{Y} \times \mathcal{M}$. Let us define*
$$\mathcal{Z}_t = \mathcal{A}_R \times \mathcal{D}_t \quad (50)$$
$$\mathcal{Q}(\mathbf{Y}_t, \mathbf{M}) = \{(\mathbf{A}, \mathbf{dM}) \in \mathcal{Z}_t|$$
$$\nabla_{(\mathbf{A}, \mathbf{dM})}f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \mathbf{0}\} \quad (51)$$
$$(\mathbf{A}_t^*, \mathbf{dM}_t^*) \in \mathcal{Q}(\mathbf{Y}_t, \mathbf{M}) \quad (52)$$
$$v(\mathbf{Y}_t, \mathbf{M}) = f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}_t^*, \mathbf{dM}_t^*). \quad (53)$$
Then
1) the function v is continuously differentiable with respect to $\mathbf{M}$ and $\nabla_{\mathbf{M}}v(\mathbf{Y}_t, \mathbf{M}) = \nabla_{\mathbf{M}}f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}_t^*, \mathbf{dM}_t^*)$;
2) g defined in (6) is continuously differentiable and $\nabla_{\mathbf{M}}g(\mathbf{M}) = \mathbb{E}_{\mathbf{Y}_t}[\nabla_{\mathbf{M}}z(\mathbf{Y}_t, \mathbf{M})]$;
3) $\nabla_{\mathbf{M}}g$ is Lipschitz continuous on $\mathcal{M}$.

Proof. The existence of local minima of $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot)$ on $\mathcal{Z}_t$ follows from the continuity of $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot)$ and the compactness of $\mathcal{Z}_t$. This ensures the non-emptiness of $\mathcal{Q}(\mathbf{Y}_t, \mathbf{M})$ and justifies the definition of $(\mathbf{A}_t^*, \mathbf{dM}_t^*)$.

Furthermore, Assumption 3 requires the invertibility of the Hessian matrix $\mathbf{H}_{(\mathbf{A}, \mathbf{dM})}f$ at the point $(\mathbf{Y}_t, \mathbf{M}, (\mathbf{A}_t^*, \mathbf{dM}_t^*))$. The first statement then follows from the implicit function theorem [29, Theorem 5.9 p.19]: there exist two open subsets $V \subset \mathcal{M}$, $W \subset \mathcal{Z}_t$ and a continuously differentiable function $\varphi : V \longrightarrow W$ such that

(i) $(\mathbf{M}, (\mathbf{A}_t^*, \mathbf{dM}_t^*)) \in V \times W \subset \mathcal{M} \times \mathcal{Z}_t$;
(ii) for all $(\tilde{\mathbf{M}}, (\mathbf{A}, \mathbf{dM})) \in V \times W$, we have
$$[\nabla_{(\mathbf{A}, \mathbf{dM})}f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}, \mathbf{dM}) = \mathbf{0}]$$
$$\Rightarrow [(\mathbf{A}, \mathbf{dM}) = \varphi(\tilde{\mathbf{M}})]; \quad (54)$$
(iii) for all $\tilde{\mathbf{M}} \in V$,
$$\frac{\partial\varphi}{\partial\mathbf{M}}(\tilde{\mathbf{M}}) = -\mathbf{H}_{(\mathbf{A}, \mathbf{dM})}^{-1}f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}}))$$
$$\frac{\partial f}{\partial\mathbf{M}\partial(\mathbf{A}, \mathbf{dM})}(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})). \quad (55)$$

In particular, $(\mathbf{M}, (\mathbf{A}_t^*, \mathbf{dM}_t^*)) \in V \times W$ satisfies (54). Then, taking the derivative of $v(\mathbf{Y}_t, \cdot)$ in $\mathbf{M}$ leads to
$$\frac{\partial v}{\partial\mathbf{M}}(\mathbf{Y}_t, \mathbf{M}) = \underbrace{\frac{\partial f}{\partial(\mathbf{A}, \mathbf{dM})}(\mathbf{Y}_t, \mathbf{M}, \varphi(\mathbf{M}))}_{=\mathbf{0}\text{ since }\varphi(\mathbf{M})\in\mathcal{Q}(\mathbf{Y}_t, \mathbf{M})}\frac{\partial\varphi}{\partial\mathbf{M}}(\mathbf{M})$$
$$+ \frac{\partial f}{\partial\mathbf{M}}(\mathbf{Y}_t, \mathbf{M}, \varphi(\mathbf{M})) \quad (56)$$

The second statement follows from the continuous differentiability of $z(\mathbf{Y}_t, \cdot)$, defined on a compact set.

We finally observe that $\|\mathbf{A}_t^*\|_{\mathrm{F}}$ and $\|\mathbf{dM}_t^*\|_{\mathrm{F}}$ are respectively bounded by a constant independent from $\mathbf{Y}_t$ (since $(\mathbf{A}_t^*, \mathbf{dM}_t^*) \in \mathcal{A}_R \times \mathcal{D}_t$). This observation, combined with the expression of $\nabla_{\mathbf{M}}f$ and the compactness of $\mathcal{M}$, leads to the third claim. $\qquad\square$

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.

[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] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4904–4917, Dec. 2015.

[6] 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.

[7] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[8] A. Halimi, N. Dobigeon, J.-Y. Tourneret, S. McLaughlin, and P. Honeine, "Unmixing hyperspectral images accounting for temporal and spatial endmember variability," in *Proc. European Signal Process. Conf. (EU-SIPCO)*, Nice, France, 2015, pp. 1686–1690.

[9] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," Oct. 2015, arXiv preprint. [Online]. Available: http://arxiv.org/pdf/1510.04238v1.pdf

[10] D. Ralph and H. Xu, "Convergence of stationary points of sample average two-stage stochastic programs: A generalized equation approach," *Mathematics of Operations Research*, vol. 36, no. 3, pp. 568–592, Aug. 2011.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learning Research*, vol. 11, pp. 19–60, Jan. 2010.

[12] 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.

[13] 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.

[14] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Reykjavik, Iceland, June 2010.

[15] 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.

[16] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[17] 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.

[18] 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.

[19] 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.

[20] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projection onto the ℓ_1-ball for learning in high dimensions," in *Proc. Int. Conf. Machine Learning (ICML)*, Helsinki, Finland, 2008.

[21] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Program., Ser. A*, pp. 1–11, Sept. 2015.

[22] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in hilbert spaces," in *Advances in Order Restricted Statistical Inference.* Springer New York, 1986, pp. 28–47.

[23] S. Becker, E. J. Candès, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Prog. Comp.*, vol. 3, no. 3, pp. 165–218, Sept. 2011.

[24] R. Heylen, A. Atker, and P. Scheunders, "On using projection onto convex sets for solving the hyperspectral unmixing problem," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1522–1526, 2013.

[25] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Online unmixing of multitemporal hyperspectral images accounting for spectral variability – complementary results," University of Toulouse, IRIT/INP-ENSEEIHT, France, Tech. Rep., Oct. 2015. [Online]. Available: http://thouvenin.perso.enseeiht.fr/papers/Thouvenin_TR_2015.pdf

[26] 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.

[27] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J.-Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using an eigen-gap approach," *IEEE Trans. Geosci. Remote Sens.*, 2016, to appear. [Online]. Available: http://arxiv.org/pdf/1501.05552v1.pdf

[28] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[29] S. Lang, *Fundamentals of differential geometry*, 1st ed., ser. Graduate Texts in Mathematics. Springer-Verlag New York, 1999, vol. 191.

Pierre-Antoine Thouvenin (S'15) 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 research is currently focused on hyperspectral unmixing and variability modeling in hyperspectral imagery.

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).