Modeling spatial and temporal variabilities in hyperspectral image unmixing

Pierre-Antoine Thouvenin

To cite this version:

Pierre-Antoine Thouvenin. Modeling spatial and temporal variabilities in hyperspectral image unmixing. Graphics [cs.GR]. Institut National Polytechnique de Toulouse - INPT, 2017. English. NNT : 2017INPT0068 . tel-01711509v2

HAL Id: tel-01711509
https://theses.hal.science/tel-01711509v2

Submitted on 29 Sep 2023

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Signal, Image, Acoustique et Optimisation

Présentée et soutenue par :

M. PIERRE-ANTOINE THOUVENIN
le mardi 17 octobre 2017

Titre :

Modeling spatial and temporal variabilities in hyperspectral image unmixing

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. NICOLAS DOBIGEON
M. JEAN-YVES TOURNERET

Rapporteurs :

M. JEAN-CHRISTOPHE PESQUET, UNIVERSITE PARIS 5
M. JÉRÔME IDIER, CNRS

Membre(s) du jury :

M. CHRISTIAN JUTTEN, UNIVERSITE JOSEPH FOURIER, Président
M. MARIO FIGUEIREDO, INSTITUTO SUPERIOR TECNICO LISBONNE, Membre
Mme VÉRONIQUE SERFATY, DGA, Membre
M. STEPHANE MAY, CENTRE NATIONAL D'ETUDES SPATIALES CNES, Membre

Remerciements

Entre figure imposée et démarche spontanée, la rédaction des remerciements reste, à de nombreux égards, un exercice complexe et délicat. Complexe, en ce qu'il comporte de contraintes stylistiques, entre considérations esthétiques, exigence de concision et souci d'exhaustivité. Délicat, dans la mesure où cet exercice nous confronte à la dimension humaine de toute entreprise scientifique, dont la thèse constitue une première expérience représentative. À ce titre, l'inévitable faiblesse du langage, résultant de l'intervalle entre les mots et les choses, ne pourra donner qu'une faible idée de ma reconnaissance envers celles et ceux qui m'ont accompagné sur ce chemin.

D'un point de vue scientifique, je souhaiterais en premier lieu exprimer toute ma gratitude envers mes directeurs de thèse, Nicolas DOBIGEON et Jean-Yves TOURNERET, dont l'entousiasme scientifique communicatif, l'expérience et les conseils avisés ont illuminé ces trois dernières années. Après m'avoir initié aux arcanes du traitement du signal en tant qu'élève ingénieur, puis m'en avoir révélé la dimension ludique lors de mon stage de master, leur enseignement et leur exemple ont été de profondes sources d'inspiration qui m'auront guidé jusqu'à la conclusion de cette thèse. J'aimerais également remercier le Pr. Christian JUTTEN pour avoir accepté de présider le jury, les professeurs Jérôme IDIER et Jean-Christophe PESQUET pour avoir évalué ce travail en tant que rapporteurs, ainsi que le Dr. Stéphane MAY, le Pr. Màrio FIGUEIREDO et Mme. Véronique SERFATY pour avoir participé au jury en tant qu'examinateurs. Leurs commentaires, lors de la phase d'évaluation du manuscrit et à l'issue de la soutenance, ont enrichi la réflexion initialement engagée avec mes directeurs, tout en ouvrant de nouvelles perspectives de travail.

Par-delà l'expérience scientifique, c'est notamment l'expérience humaine qu'il convient de remettre en perspective. Aux antipodes d'une science sans conscience, cette thèse a consitué une opportunité unique de cotoyer des personnalités fortes, que ce soit lors des conférences, des réunions, ou plus simplement autour d'un café au labo. À cet égard, le contact avec mes directeurs et les autres membres de l'équipe SC aura été riche en enseignements. En particulier, la bonne humeur des (post-)doctorants de l'équipe a instauré une atmosphère de travail et d'échange idéale lors de ces nombreuses heures passées à l'ENSEEIHT : merci à Qi, Ningning, Sébastien, Nora, Tarik, Jean-Adrien, Matthieu, Vinicius, Yanna, Adrien, Etienne, Tatsumi, Olivier, Louis et Dylan. Je tiens également à remercier les secrétaires, Annabelle, Sylvie et Isabelle, pour leur soutien et leur patience face aux nombreuses questions administratives que j'ai pu leur poser en cours de thèse.

Je souhaiterais enfin conclure ces remerciements par une pensée pour les membres de ma famille, qu'ils aient pu ou non assister à la conclusion de cette aventure, ainsi que les amis avec lesquels le contact s'est maintenu, intact, en dépit de l'emploi du temps chargé des uns et des autres.

Édimbourg, le 20 Octobre 2017.

> “ *Et c'est en somme une façon comme une autre de résoudre le problème de l'existence, qu'approcher suffisamment les choses et les personnes qui nous ont paru de loin belles et mystérieuses, pour nous rendre compte qu'elles sont sans mystère et sans beauté.* ”
>
> ---
>
> Marcel Proust, *À l'ombre des jeunes filles en fleurs, Nom de pays : le pays*, 1918

> “ *Il y a une vision du monde proustienne. Elle se définit d'abord par ce qu'elle exclut : ni matière brute, ni esprit volontaire. Ni physique, ni philosophie. La philosophie suppose des énoncés directs et des significations explicites, issus d'un esprit qui veut le vrai. La physique suppose une matière objective et non ambiguë, soumise aux conditions du réel. Nous avons tort de croire aux faits, il n'y a que des signes. Nous avons tort de croire à la vérité, il n'y a que des interprétations.* ”
>
> ---
>
> Gilles Deleuze, *Proust et les signes, Le pluralisme*, 1964

Résumé

Acquises dans plusieurs centaines de bandes spectrales contiguës, les images hyperspectrales permettent d'analyser finement la composition d'une scène observée. En raison de la résolution spatiale limitée des capteurs utilisés, le spectre d'un pixel d'une image hyperspectrale résulte de la composition de plusieurs signatures associées à des matériaux distincts. À ce titre, le démélange d'images hyperspectrales vise à estimer les signatures des différents matériaux observés ainsi que leur proportion dans chacun des pixels de l'image.

Pour cette analyse, il est d'usage de considérer qu'une signature spectrale unique permet de décrire un matériau donné, ce qui est généralement intrinsèque au modèle de mélange choisi. Toutefois, la signature d'un matériau présente en pratique une variabilité spectrale qui peut être significative d'une image à une autre, voire au sein d'une même image. De nombreux paramètres peuvent en être cause, tels que les conditions d'acquisitions (*e.g.*, conditions d'illumination locales), la déclivité de la scène observée ou des interactions complexes entre la lumière incidente et les éléments observés. À défaut d'être prises en compte, ces sources de variabilité perturbent fortement les signatures extraites, tant en termes d'amplitude que de forme. De ce fait, des erreurs d'estimation peuvent apparaître, qui sont d'autant plus importantes dans le cas de procédures de démélange non-supervisées. Le but de cette thèse consiste ainsi à proposer de nouvelles méthodes de démélange pour prendre en compte efficacement ce phénomène.

Nous introduisons dans un premier temps un modèle de démélange original visant à prendre explicitement en compte la variabilité spatiale des spectres purs. Les paramètres de ce modèle sont estimés à l'aide d'un algorithme d'optimisation sous contraintes. Toutefois, ce modèle s'avère sensible à la présence de variations spectrales abruptes, telles que causées par la présence de données aberrantes ou l'apparition d'un nouveau matériau lors de l'analyse d'images hyperspectrales multi-temporelles. Pour pallier ce problème, nous introduisons une procédure de démélange robuste adaptée à l'analyse d'images multi-temporelles de taille modérée. Compte tenu de la dimension importante des données étudiées, notamment dans le cas d'images multi-temporelles, nous avons par ailleurs étudié une stratégie d'estimation en ligne des différents paramètres du modèle de mélange proposé. Enfin, ce travail se conclut par l'étude d'une procédure d'estimation distribuée asynchrone, adaptée au démélange d'un grand nombre d'images hyperspectrales acquises sur une même scène à différents instants.

Mots clés : Imagerie hyperspectrale, séparation aveugle de sources, optimisation non-convexe, optimisation stochastique, méthodes de Monte-Carlo par chaîne de Markov.

Abstract

Acquired in hundreds of contiguous spectral bands, hyperspectral (HS) images have received an increasing interest due to the significant spectral information they convey about the materials present in a given scene. However, the limited spatial resolution of hyperspectral sensors implies that the observations are mixtures of multiple signatures corresponding to distinct materials. Hyperspectral unmixing is aimed at identifying the reference spectral signatures composing the data – referred to as *endmembers* – and their relative proportion in each pixel according to a predefined mixture model. In this context, a given material is commonly assumed to be represented by a single spectral signature. This assumption shows a first limitation, since endmembers may vary locally within a single image, or from an image to another due to varying acquisition conditions, such as declivity and possibly complex interactions between the incident light and the observed materials. Unless properly accounted for, spectral variability can have a significant impact on the shape and the amplitude of the acquired signatures, thus inducing possibly significant estimation errors during the unmixing process. A second limitation results from the significant size of HS data, which may preclude the use of batch estimation procedures commonly used in the literature, i.e., techniques exploiting all the available data at once. Such computational considerations notably become prominent to characterize endmember variability in multi-temporal HS (MTHS) images, i.e., sequences of HS images acquired over the same area at different time instants.

The main objective of this thesis consists in introducing new models and unmixing procedures to account for spatial and temporal endmember variability. Endmember variability is addressed by considering an explicit variability model reminiscent of the total least squares problem, and later extended to account for time-varying signatures. The variability is first estimated using an unsupervised deterministic optimization procedure based on the Alternating Direction Method of Multipliers (ADMM). Given the sensitivity of this approach to abrupt spectral variations, a robust model formulated within a Bayesian framework is introduced. This formulation enables smooth spectral variations to be described in terms of spectral variability, and abrupt changes in terms of outliers. Finally, the computational restrictions induced by the size of the data is tackled by an online estimation algorithm. This work further investigates an asynchronous distributed estimation procedure to estimate the parameters of the proposed models.

Keywords: Hyperspectral imagery, blind source separation, non-convex optimization, stochastic optimization, Markov chain Monte-Carlo (MCMC) methods.

Contents

List of Figures

List of Tables

Introduction (in French)

Contexte et problématique de la thèse

L'imagerie hyperspectrale a fait l'objet d'un grand nombre de travaux de recherche au cours des dernières années, motivés essentiellement par l'information spectrale que ces images apportent quant aux matériaux présents dans une scène observée. En pratique, les spectres acquis sont souvent constitués d'un mélange de signatures associées à des matériaux distincts, en raison de la résolution spatiale limitée des capteurs utilisés. À cet égard, le démélange d'images hyperspectrales a pour but d'identifier ces signatures de référence (spectres purs), ainsi que de quantifier leur proportion en chacun des pixels suivant un modèle de mélange prédéfini. En général, le choix d'un modèle de mélange reflète une connaissance *a priori* de l'analyste quant à la nature des phénomènes physiques pouvant affecter les observations (*e.g.*, déclivité de la scène, réflections multiples) [Bio+12; Dob+14b; HPG14]. Il est alors d'usage de supposer qu'une signature spectrale caractérise un matériau donné de façon univoque. Toutefois, la validité de cette hypothèse est mise en défaut en présence de fortes variations des conditions d'acquisition, telles que l'évolution naturelle du milieu observé ou une variation locale des conditions d'illumination. Ces variations ont un impact important sur la forme et l'amplitude des signatures acquises [ZH14], ce qui résulte en une forte variabilité des spectres purs.

Jusqu'à présent, de nombreux travaux se sont attachés à prendre en compte ce phénomène lors de l'analyse d'une image donnée. À ce titre, deux grandes classes d'approches ont été proposées dans un cadre soit statistique, soit déterministe. Tandis que la première classe de méthodes représente les spectres purs par la réalisation de vecteurs aléatoires [Ech+10; Du+14; HDT15], la seconde s'appuie sur l'utilisation d'un dictionnaire de signatures – disponible *a priori* ou extrait des données – pour appréhender la variabilité spectrale [JWZ10; Som+12]. Enfin, une autre méthode récente suppose que la variabilité résulte exclusivement de facteurs d'échelle spatialement variables [Veg+14; Dru+16]. Tandis que la variabilité spectrale au sein d'une seule image a déjà fait l'objet d'un grand nombre d'études, ce phénomène n'a été que peu étudié lors de l'analyse d'images hyperspectrales multi-temporelles. Par ailleurs, les méthodes de démélange proposées dans ce contexte [HCJ16; Hal+15] restent sensibles aux variations spectrales abruptes, fréquemment observées dans les données réelles disponibles.

Dans ce contexte, cette thèse vise à développer de nouvelles techniques de démélange afin de mieux représenter la variabilité spatiale et temporelle des acquisitions. Le travail présenté dans ce manuscrit a été mené au sein de l'équipe Signal et Communications de l'Institut de Recherche en Informatique de Toulouse (IRIT), équipe au sein de laquelle l'analyse d'images hyperspectrales a déjà motivé plusieurs travaux de doctorats [Dob07; Ech10; Alt13; Wei15]. Cette thèse a été intégralement financée par la Direction Générale de l'Armement, relevant du ministère français de la défense.

Structure du manuscrit

Le premier chapitre de ce manuscrit présente un modèle de mélange original visant à prendre en compte explicitement la variabilité spatiale des signatures au sein d'une image hyperspectrale. Basé sur une formulation inspirée d'un problème de moindres carrés totaux [GL80] sous contraintes, le modèle proposé consiste à représenter les observations par une combinaison convexe de spectres purs, affectés chacun par une perturbation additive. Ces perturbations sont supposées spatialement variables afin de représenter la variabilité spatiale éventuelle des spectres purs. Sachant que la variabilité spectrale présente une contribution limitée au signal observé, l'énergie de la variabilité capturée par le modèle est restreinte *a priori*. Les paramètres de ce modèle, estimés à l'aide d'un premier algorithme d'optimisation basé sur la méthode des multiplicateurs de Lagrange (ADMM) [Boy+10], permet d'obtenir des résultats prometteurs sur données synthétiques et réelles. Les performances d'un deuxième algorithme, découlant de l'algorithme PALM (*proximal alternating linearized algorithm*) [BST13; CPR16] et disposant de garanties de convergence mieux établies que l'algorithme précédent, sont également étudiées.

Inspiré par les travaux [Hal+15; AMH15], le deuxième chapitre est dédié à l'estimation conjointe de la variation temporelle des spectres purs (décrite en termes de variabilité spectrale), ainsi que des variations éventuellement brutales des signatures spectrales au cours du temps. L'approche proposée concerne plus particulièrement l'étude d'images acquises pour une même scène à différents instants, où certains matériaux sont de fait observés dans plusieurs des images acquises. Formulé dans un cadre bayésien, le modèle proposé vise à réduire la sensibilité à des variations spectrales abruptes des méthodes précédemment proposées dans ce contexte [HCJ16; Hal+15], tout en exploitant la donnée temporelle à disposition. Les paramètres de ce modèle sont estimés à l'aide d'un algorithme de Monte-Carlo par chaîne de Markov. Les résultats obtenus montrent l'intérêt d'une analyse conjointe de plusieurs images hyperspectrales, comparée à une analyse individuelle de ces mêmes images.

Par ailleurs, la taille des données considérées peut compromettre la mise en œuvre de méthodes d'estimation conjointes, i.e., qui exploitent simultanément la totalité des données disponibles. Dans cette perspective, le troisième chapitre s'intéresse à un algorithme d'estimation en ligne adapté à l'analyse d'images hyperspectrales multi-temporelles. Formulé sous la forme d'un problème d'optimisation stochastique avec recours [RX11; Mai+10], le problème de démélange en présence de variabilité est abordé à l'aide d'un algorithme inspiré de celui proposé dans [Mai+10].

Le quatrième chapitre envisage finalement le développement d'un algorithme d'optimisation distribué asynchrone, fondé sur des contributions récentes en analyse non-convexe [PR15; Sra+16; Cha+16; Scu+17; Dav16; HY16; Can+16; CE16; Pen+16]. Offrant un point de vue complémentaire à celui proposé dans le chapitre précédent, l'intérêt d'une procédure d'optimisation distribuée asynchrone, par rapport à une version synchrone du même algorithme, est illustré sur des données synthétiques et réelles.

Principales contributions

Chapter 1. Un modèle explicite de variabilité spatiale est proposé dans le cadre du démélange d'images hyperspectrales. Un premier algorithme d'optimisation, qui consiste à résoudre une séquence de sous-problèmes strictement convexes par la méthode des multiplicateurs de Lagrange (ADMM) dans un algorithme de descente par coordonnées, permet d'estimer les paramètres du modèle. Les performances de l'algorithme PALM [BST13; CPR16] sont également évaluées dans ce contexte. Ces deux algorithmes permettent d'obtenir des résultas de reconstruction compétitifs, tout en donnant une représentation satisfaisante de la variabilité.

Chapter 2. Un modèle bayésien hiérarchique est proposé pour le démélange robuste d'images hyperspectrales multi-temporelles. Ce modèle de mélange rend compte d'une variation temporelle modérée des spectres purs – décrite en termes de variabilité – et de changements spectraux abrupts – interprétés en termes de données aberrantes. Le modèle proposé s'avère particulièrement pertinent pour l'analyse de données réelles, notamment lorsque l'intervalle de temps entre deux acquisitions est important. Dans un tel cas, l'apparition d'un matériau non présent dans la scène de référence induit par exemple de telles variations.

Chapter 3. Une stratégie de démélange en ligne est proposée pour analyser une collection d'images hyperspectrales, acquises au-dessus d'une même scène d'intérêt à différents instants. Basé sur une variante du modèle introduit dans le permier chapitre, le problème de démélange est formulé sous la forme d'un problème d'optimisation stochastique avec recours. Les résultats obtenus sur données synthétiques et réelles montrent l'intérêt d'utiliser plusieurs images hyperspectrales pour améliorer les résultats du démélange.

Chapter 4. La mise en œuvre d'un algorithme distribué asynchrone, basé sur de récentes contributions en optimisation distribuée, est envisagée pour l'estimation des paramètres d'un modèle de mélange linéaire et du modèle introduit dans le premier chapitre. Les simulations proposées visent à évaluer l'intérêt de l'asynchronie pour réduire le temps de calcul nécessaire à l'estimation des paramètres, en comparaison avec une version synchrone du même algorithme.

Introduction

Context and objective of the thesis

Over the past decades, hyperspectral imagery has received an increasing interest, resulting from the significant spectral information it conveys about the materials contained in the observed scene. In practice, the observations are mixtures of multiple signatures corresponding to distinct materials due to the limited spatial resolution of hyperspectral sensors. In this context, spectral unmixing consists in identifying the reference signatures composing a hyperspectral (HS) image – referred to as endmembers – and their abundance fractions in each pixel according to a predefined mixture model. The choice of a specific model generally reflects the practitioners' prior knowledge on the environmental factors possibly affecting the data (e.g., declivity, multiple reflections) [Bio+12; Dob+14b; HPG14]. Traditionally, a given material is assumed to be fully characterized by a single signature. However, this assumption presents severe limitations when the observations are affected by varying acquisition conditions (e.g., local illumination variations, natural evolution of the scene). In practice, these phenomena have a significant impact on the acquired signatures [ZH14], referred to as spectral variability.

Spectral variability within a single image has been hitherto considered in various models, either derived from a statistical or a deterministic point of view [ZH14]. The first class of methods assumes that the endmember spectra are realizations of multivariate random variables [Ech+10; Du+14; HDT15]. The second class of methods represents endmember signatures as members of spectral libraries associated with each material (bundles) [JWZ10; Som+12]. In another recent approach, the variability is assumed to exclusively result from scaling factors [Veg+14; Dru+16]. While endmember variability has been extensively considered within a single HS image, fewer works have considered temporal variability by exploiting multitemporal HS (MTHS) images. In addition, unmixing approaches previously proposed in this context [HCJ16; Hal+15] show a notable sensitivity to abrupt spectral variations, which are commonly observed in real MTHS images.

In this context, this thesis is aimed at designing new unmixing techniques to account for spectral variability, in order to improve the estimation results obtained from HS or MTHS images. The work presented in this manuscript has been carried out within the Signal and Communications group of the Institut de Recherche en Informatique de Toulouse, where several doctoral works have been previously conducted on hyperspectral image analysis [Dob07; Ech10; Alt13; Wei15]. This thesis has been funded by the Direction Générale de l'Armement, French ministry of Defense.

I.1. Linear unmixing of hyperspectral images

Whereas traditional red / green / blue and multispectral images are composed of a few spectral channels (from three to tens), hyperspectral (HS) images are acquired in hundreds of contiguous spectral bands. These images have received an increasing interest due to the significant spectral information they convey. Indeed, the spectral dimension of HS data, i.e., the different wavelengths at which the scene is observed, allow different materials to be discriminated based on their reflectance, as illustrated in Fig. I.1. However, the high spectral resolution of these images is mitigated by their lower spatial resolution in remote sensing applications. Current HS sensors used in remote sensing applications present a spatial resolution from $0.5\,\mathrm{m}$ to $20\,\mathrm{m}$. This limitation, combined with possibly complex interactions between the incident light and the observed materials, implies that the observed spectra are mixtures of several signatures corresponding to distinct materials. Spectral unmixing then consists in identifying a limited number of reference spectral signatures composing the data – referred to as *endmembers* – and their abundance fractions in each pixel according to a predefined mixture model. The choice of a specific model generally reflects any prior knowledge on the environmental factors possibly affecting the data, such as declivity or multiple reflections.

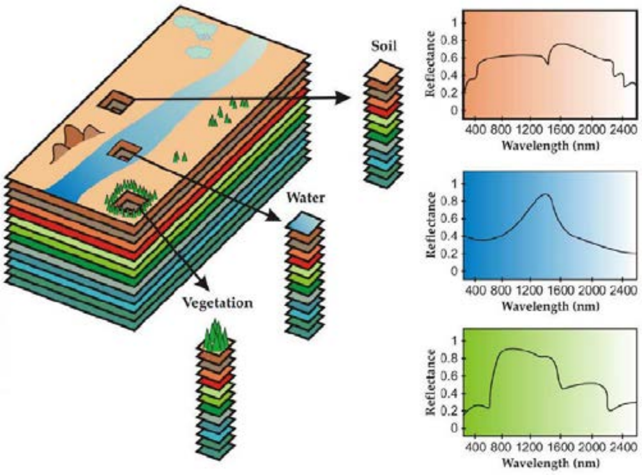


Figure I.1.: Hyperspectral unmixing: an illustration (illustration taken from [Bio+12]).

Traditionally, a linear mixing model (LMM) is adopted when the declivity of the scene and microscopic interactions between the observed materials are negligible [Bio+12]. Each spectrum associated with a pixel of the image is represented as a convex combination of R signatures $\mathbf{m}_r$ corresponding to distinct materials. More explicitly, the LMM can be expressed as

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \mathbf{b}_n \tag{I.1}$$

for $n = 1, \dots, N$, where N is the number of pixels composing the HS image of interest, R is the number of endmembers contained in the image, $\mathbf{y}_n$ denotes the nth image pixel, $\mathbf{m}_r$ is the rth endmember and $a_{r,n}$ is the proportion of the rth endmember in the nth pixel. Throughout this work, the number of endmembers R is assumed to be *a priori* known or estimated by any state-of-the-art method (e.g., [BN08; Hal+16b]). Its choice drastically alters the representation of the imaged scene, and thus remains a crucial step to the endmember identification and the subsequent abundance estimation

[BN08; Bio+12; Hal+16b]. Finally, $\mathbf{b}_n$ represents an additive noise resulting from the data acquisition and the modeling errors. The LMM can be written in matrix form as

$$\mathbf{Y} = \mathbf{M}\mathbf{A} + \mathbf{B} \tag{I.2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ is an $L \times N$ matrix containing the pixels of the image, $\mathbf{M}$ denotes an $L \times R$ matrix containing the endmembers, $\mathbf{A}$ is an $R \times N$ matrix composed of the abundance vectors $\mathbf{a}_n$, and $\mathbf{B}$ is a noise matrix. Since the analysis is generally conducted on reflectance data, the following constraints are usually considered to allow a physical interpretation of the results

$$\mathbf{A} \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}^{\mathrm{T}}\mathbf{1}_R = \mathbf{1}_N, \quad \mathbf{M} \succeq \mathbf{0}_{L,R} \tag{I.3}$$

where $\succeq$ denotes a term-wise inequality. Given the mixture model (I.2) and the constraints (I.3), the unmixing problem consists in inferring the abundances $\mathbf{A}$ and the endmembers $\mathbf{M}$. As such, spectral unmixing (SU) is an instance of a matrix factorization problem, subject to application-specific constraints.

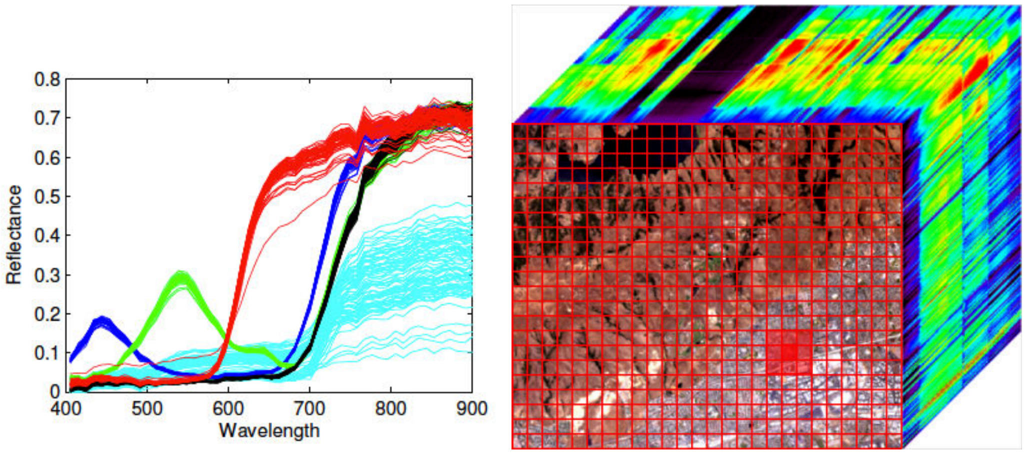
To address the limitations of the LMM occurring in specific applications, e.g., to unmix vegetated [Dob+14a] or urban areas [Meg+14], various models have been investigated to capture higher order interactions (i.e., nonlinearities) between the incident light and the observed materials. We may mention the extensively used bilinear models based on the LMM, which present an additional term whose expression is mainly derived from physical considerations [Fan+09; NB09; Hal+11; Alt+12; Meg+14; ADT14]. This residual term is expressed as a function of the endmembers and the abundance coefficients from the LMM, and explicitly determines the type of interaction captured by the model. Different approaches also investigate the use of reproducing kernel Hilbert spaces to represent the nonlinearities resulting from the interactions between different materials [CRH13; Amm+17]. Since the present work does not specifically focus on this issue, the reader is referred to [Dob+14b; HPG14] for recent reviews on this topic.

Both linear and non-linear unmixing methods rely on the implicit assumption that a given material is represented by a single spectral signature. However, this assumption may present severe limitations in practice, as detailed in the following section.

I.2. Spatial endmember variability

In practice, varying acquisition conditions, such as local illumination variations or the natural evolution of the scene, may significantly alter the shape and the amplitude of the acquired signatures [Som+11; ZH14], thus affecting the extracted endmembers. This phenomenon, illustrated in Fig. I.2(a), is generally referred to as spectral variability. In the rest of this manuscript, we will refer to spatial endmember variability whenever it results from local variations within a single HS image (see Fig. I.2(b)). This terminology is introduced to make a distinction with the problem investigated in the next section, which involves different images of the same scene acquired at different time instants.

Since the endmember signatures contained in a given image can vary spatially due to varying acquisition conditions,



(a) Spectral variability

(b) An hyperspectral image

Figure I.2.: An illustration of spectral variability (Fig. I.2(a) is taken from [Gad+13]). Each spectrum in Fig. I.2(a) corresponds to an acquisition made for a given material, where each material is represented with a different color. The signature corresponding to a given material may vary locally (Fig. I.2(b)).

resorting to a traditional LMM may result in the propagation of significant estimation errors during the SU process. Various models have been proposed in the literature to address this issue, either derived from a statistical or a deterministic point of view [ZH14; DCJ16]. More precisely, a first class of methods assumes that the endmember signatures can be considered as realizations of random vectors described by multivariate probability distributions. Within the second class of methods, several procedures consider the endmember spectra as members of spectral libraries associated with each material (bundles). These different approaches, illustrated in Fig. I.3, are further detailed in the following paragraphs.

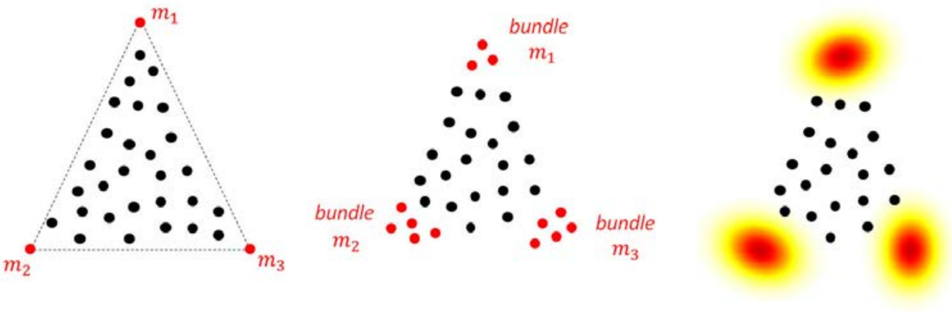


Figure I.3.: Different representations of endmember variability within the simplex enclosing the data (illustration taken from [HDT15]) [from left to right: classical LMM (single endmembers), endmember bundles and endmembers as realizations of a mutivariate probability distribution].

I.2.1. Statistical approaches

The statistical approaches previously proposed in the literature implicitly account for endmember variability by considering endmembers are realizations of multivariate probability distributions. Among these approaches, the most popular models are the normal compositional model (NCM) [Ech+10] and the beta compositional model (BCM) [Du+14].

Normal compositional model (NCM) The NCM [Ech+10] represents each pixel as a convex combination of realizations of Gaussian random vectors. This model can be expressed as

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_{r,n}, \quad \text{with} \quad \mathbf{m}_{r,n} \sim \mathcal{N}(\mathbf{m}_r, \operatorname{diag}(\boldsymbol{\sigma}_r^2)). \tag{I.4}$$

A recently proposed variant of the NCM, denoted as the generalized NCM (GNCM) [HDT15], allows spatial variability to be more explicitly represented. More precisely, this model considers band-dependent variances in the endmember prior distribution, as well as an additive noise term. The GNCM can be written

$$\begin{aligned} \mathbf{y}_n &= \sum_{r=1}^{R} a_{r,n}\mathbf{m}_{r,n} + \mathbf{b}_n, \\ \mathbf{m}_{r,n} &\sim \mathcal{N}(\mathbf{m}_r, \operatorname{diag}(\boldsymbol{\sigma}_r^2)), \quad \mathbf{b}_n \sim \mathcal{N}(\mathbf{0}_L, \psi_n^2\mathbf{I}_L). \end{aligned} \tag{I.5}$$

In practice, endmember variability is expected to represent a more significant fraction of the observed spectra than the noise [HDT15]. The variances ψ_n^2 are thus assigned an exponential prior whose parameter has a large value, which ensures the variances ψ_n^2 are of limited amplitude. When compared to the original NCM, this formulation allows spectral variations to be captured for each element of interest in each spectral band.

Beta compositional model (BCM) For some materials, such as the grass, the beta distribution has been observed to provide a better fit of the reflectance histogram than the Gaussian distribution [Du+14]. A mixture model involving realizations of beta distributed vectors has been consequently proposed in [Du+14]

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r, \quad \text{with} \quad m_{\ell,r} \sim \mathcal{B}e(\alpha_{\ell,r}, \beta_{\ell,r}) \tag{I.6}$$

where $\mathcal{B}e$ denotes the beta distribution, and the parameters $(\alpha_{\ell,r}, \beta_{\ell,r})$ are learnt from the data prior to the unmixing procedure.

I.2.2. Deterministic approaches

Deterministic approaches proposed in the literature essentially rely on the use of an appropriate endmember dictionary, referred to as endmember bundle. Composed of multiple candidate endmembers for each material, the dictionary is either *a priori* available (e.g., composed of signatures acquired in a controlled environment), or directly extracted from the data [Rob+98; Som+12]. Other recent approaches consider extensions of the LMM to capture potential sources of variability within the image under study [Dru+16].

Endmember bundle extraction. Endmember bundles is a widely used solution aimed at exploiting the diversity of different signatures to better represent a given material [Rob+98]. Extracting a representative dictionary is thus a preliminary step before applying these unmixing techniques. An automatic extraction procedure, namely the automated endmember bundle (AEB) [Som+12], has been notably proposed for this purpose. This technique consists in decomposing the image into randomly selected subsets of pixels, to which an endmember extraction algorithm (EEA) is individually applied. A clustering algorithm is finally used on the collection of the extracted signatures to recover meaningful material classes. Another recent contribution in this context includes [Uez+16a].

Unmixing techniques based on endmember bundles. Once a suitable dictionary is available, an extensive number of methods has been proposed to estimate meaningful abundance coefficients. For instance, the Fisher discriminant null space (FDNS) [JWZ10] is aimed at estimating a transformation projection matrix to project the hyperspectral data into a space minimizing the variability impact. More precisely, this method is aimed at estimating a subspace in which the intra-class variability of the bundle is minimized, whereas the inter-class variability is maximized. The abundance coefficients may then be estimated by any state-of-the-art optimization technique, such as [BF10] in the context of HS unmixing. Other methods include the use of sparse unmixing [IBP11; Mey+16], Gaussian processes [Uez+16b; Uez+16c] or endmember selection procedures [Rob+98; Hey+16]. Given the significant number of methods proposed in this context, the reader is referred to the recent reviews [Som+11; ZH14; DCJ16] for further details. One may nevertheless note that the performance of these methods is essentially conditioned by the quality of the dictionary.

Explicit variability models. A recent class of approaches, including the model introduced in chapter 1, consists in introducing an explicit variability model whose additional parameters are intended to capture the observed variability. An alternative model has been proposed by [Veg+14; Dru+16], in which the variability is assumed to exclusively result from scaling factors, based on a simplification of the Hapke model [Hap81; Hap93] and on considerations on the influence of the geometry of the scene [NB05b; NB05a]. The model is expressed in [Dru+16] as

$$\mathbf{Y} = \mathbf{M}(\boldsymbol{\Psi} \odot \mathbf{A}) + \mathbf{B} \tag{I.7}$$

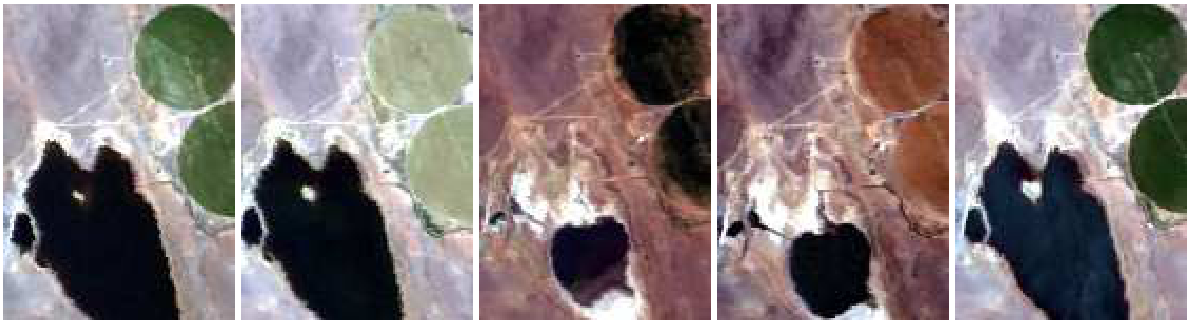
where $\odot$ denotes the Hadamard (term-wise) product. The matrix $\boldsymbol{\Psi} \in \mathbb{R}^{R \times N}$ represents the scaling factors locally affecting the endmembers composing the data. The authors then proposed to estimate the parameters of this model by resorting to an algorithm similar to the one proposed in chapter 1. The scaling indeterminacy introduced by the presence of $\boldsymbol{\Psi}$ is partly addressed in the estimation algorithm proposed in [Dru+16] by minimizing the cost function

$$\mathcal{J}(\mathbf{A}, \underline{\mathbf{M}}, \boldsymbol{\Psi}) = \frac{1}{2} \sum_{n=1}^{N} \left(\|\mathbf{y}_n - \mathbf{M}_n \mathbf{a}_n\| + \lambda \, \|\mathbf{M}_n - \mathbf{M}_0 \boldsymbol{\psi}_n\|_{\mathrm{F}}^2 \right) + g(\mathbf{A}) + h(\boldsymbol{\Psi}) \tag{I.8}$$

where $\underline{\mathbf{M}}$ denotes the collection of the matrices $\mathbf{M}_n$, $\boldsymbol{\psi}_n$ denotes the nth column of $\boldsymbol{\Psi}$, $\mathbf{M}_0$ denotes a reference endmember matrix and g, h are appropriate penalization functions.

I.3. Towards the unmixing of multi-temporal hyperspectral images

While spatial endmember variability has hitherto been extensively considered in the literature, fewer works have considered temporal variability by exploiting the possibilities offered by multitemporal HS (MTHS) images. For unmixing applications, such data have been notably used in the context of change detection between two HS images [EP15; Liu+16], or in the perspective of improving unmixing results by incorporating data acquired at different time instants by possibly different sensors [YZP17]. Indeed, the use of such data is particularly appealing to improve the unmixing results, especially when the observed materials are not jointly well represented in a single image, but in different images of the scene of interest (see Fig. I.4). The results obtained by such unmixing strategies are consequently expected to outperfom those relying on an individual analysis of each image [Goe+13]. As for temporal variability, the recent contributions [Hal+15; HCJ16] are among the first to specifically address this issue in hyperspectral unmixing problems. In a slightly different context (i.e., multi-angular HS images), we may also mention the algorithms recently proposed in [Veg+15a; Veg+15b]. However, the use of MTHS data raises two main practical limitations.



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

Figure I.4.: An example of a sequence of real HS images, acquired at different time instants over the same area.

- **Abrupt spectral variations:** even though the unmixing approaches [Hal+15; HCJ16] enable smooth variations of some of the mixture parameters to be accounted for, they remain sensitive to the presence of abrupt spectral variations. Such variations are commonly observed in sequences of real hyperspectral data, either due to the presence of outliers or to possibly significant time intervals between consecutive acquisitions. In the latter situation, the natural evolution of the scene may induce significant changes, e.g., when water or vegetation is present in the observed scene, as illustrated in Fig. I.4. Unless specifically accounted for, these variations have a significant impact on the recovered endmembers. This issue is specifically addressed in Chapter 2.

- **Size of the data:** HS unmixing of a significant number of images, or several large images, may preclude the use of batch estimation procedures as in [Hal+15; HCJ16; YZP17] 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, Chapter 3 will be focused on the design of an online HS unmixing method accounting for temporal variability. Furthermore, the development of unmixing procedures based on [CPR16] will be shown to be particularly relevant for MTHS unmixing. From a complementary point of view, the development of a distributed asynchronous unmixing algorithm is investigated in Chapter 4.

I.4. Measures of unmixing quality

The performance of the unmixing strategies compared in this manuscript is evaluated on synthetic data in terms of the quality of the recovered abundance, endmember and variability estimates, as well as the reconstruction error. More explicitly, quality assessment is conducted in terms of the following criteria.

(i) **Average spectral angle mapper (aSAM).** The endmember estimation is evaluated through the average spectral angle mapper (aSAM), which is particularly relevant to evaluate the similarity between the shape of the true and the recovered endmembers

$$\text{aSAM}(\mathbf{M}) = \frac{1}{R}\sum_{r=1}^{R} \arccos\left(\frac{\mathbf{m}_r^{\text{T}}\widehat{\mathbf{m}}_r}{\|\mathbf{m}_r\|_2\|\widehat{\mathbf{m}}_r\|_2}\right); \tag{I.9}$$

(ii) **Global mean square errors (GMSEs).** The abundance and variability estimations are evaluated through the global mean square errors (GMSEs)

$$\text{GMSE}(\mathbf{A}) = \frac{1}{RN}\|\mathbf{A} - \widehat{\mathbf{A}}\|_{\text{F}}^2 \tag{I.10}$$

$$\text{GMSE}(\mathbf{dM}) = \frac{1}{LR}\|\mathbf{dM} - \widehat{\mathbf{dM}}\|_{\text{F}}^2; \tag{I.11}$$

(iii) **Reconstruction error (RE).** The accuracy of the reconstructed data is assessed in terms of the quadratic reconstruction error (RE)

$$\text{RE} = \frac{1}{LN}\|\mathbf{Y} - \widehat{\mathbf{Y}}\|_{\text{F}}^2 \tag{I.12}$$

where $\widehat{\mathbf{Y}}$ is the matrix composed of the pixels reconstructed with the estimated parameters.

The reconstruction error RE must be carefully interpreted depending on the context of the study, since this criterion does not provide a sufficient information to assess the unmixing performance. Indeed, a low value of the RE may correspond to a trivial representation of the data (i.e., when the noise is captured by the unmixing procedure), and significantly depends on the number of degrees of freedom introduced in the model. As such, this criterion must be systematically interpreted in conjunction with other sources of information available on the data:

- for synthetic data, the RE provides a complementary information in terms of the algorithm convergence when coupled with the RMSEs computed for the different parameters. It also provides a useful indication against the risk of overfitting. Ideally, the performance of the algorithms should be exclusively compared in terms of the RMSEs for a given RE;
- for real data, the interpretation of the RE must be conducted in association with a semantic analysis of the results (e.g., consistency of the extracted endmembers and abundances, spatial distribution of the estimated variability),

especially in the absence of ground truth.

In both cases, the RE is not sufficient to properly quantify the unmixing performance of an algorithm, and should be systematically related to complementary sources of information available on the data under study.

Structure of the manuscript

Chapter 1 introduces an explicit model to account for spatial endmember variability, i.e., locally occurring within a single image. Reminiscent of the total least squares problem [GL80], the observations are modeled as a linear mixture of perturbed endmembers, in the sense that they are affected by a spatially varying additive perturbation to account for endmember variability. The parameters of this model are estimated using an optimization algorithm based on the alternating direction method of multipliers (ADMM) [Boy+10], showing promising results. An alternative estimation strategy based on the proximal alternating linearized minimization (PALM) [BST13] algorithm is also reported. 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.

Chapter 2 introduces a robust unmixing model inspired by [Hal+15; AMH15], specifically designed to address multitemporal hyperspectral images of moderate spatial and temporal dimensions. More specifically, HS images are likely to simultaneously exhibit moderate endmember variability and abrupt spectral changes either due to outliers or to significant time intervals between consecutive acquisitions. Unless properly accounted for, these two perturbations can significantly affect the unmixing process. In this context, we propose a new unmixing model for multitemporal hyperspectral images, accounting for smooth temporal variations – construed as spectral variability – and abrupt spectral changes interpreted as outliers. The parameters involved in the proposed hierarchical Bayesian model are inferred using a Markov chain Monte-Carlo (MCMC) method, which enables the posterior of interest to be sampled and Bayesian estimators to be approximated. A comparison with unmixing techniques from the literature illustrates the interest of the proposed approach on synthetic and real data.

Chapter 3 investigates an online unmixing algorithm to address the estimation of temporal endmember variability potentially affecting multitemporal hyperspectral images, i.e., sequence of images acquired over the same area at different time instants. In this context, the observations are likely to be composed of the same materials. In addition, the possibly significant size of the data under study may preclude the use of traditional batch estimation techniques to jointly estimate the mixture parameters. The online hyperspectral unmixing problem is formulated as a two-stage stochastic program [RX11; Mai+10], which can be solved using a stochastic approximation. The performance of the proposed method is evaluated on synthetic and real data. A comparison with different unmixing algorithms finally illustrates the interest of the proposed strategy. The results obtained in this chapter suggest that this analysis enables the endmembers of the scene to be tracked and the corresponding endmember variability to be consistently characterized.

Chapter 4 focuses on the applicability of asynchronous distributed optimization techniques to estimate the mixture parameters involved in the model considered in the preceding chapters. Complementary to online estimation algorithms from a computational point of view, distributed optimization techniques appear as a relevant alternative given the increase in the number of image acquisitions, and the fact that the data may be distributed between different computational nodes. In this context, we propose an asynchronous distributed unmixing algorithm based on recent advances in asynchronous non-convex optimization [PR15; Sra+16; Cha+16; Scu+17; Dav16; HY16; Can+16]. The gain brought by the asynchronous algorithm, when compared to its synchronous counterpart, is illustrated on synthetic and real data.

Main contributions

Chapter 1. The contribution of this chapter lies in the introduction of an explicit parametric variability model. A block coordinate descent algorithm is first proposed to estimate the parameters of interest, which consists in solving strongly convex subproblems by ADMM. The proposed approach yields competitive reconstruction performance and consistent variability estimates on both synthetic and real data. The performance of the proximal algorithm [BST13; CPR16] is finally evaluated for variability estimation.

Chapter 2. A hierarchical Bayesian model is introduced to analyze multitemporal hyperspectral images. This contribution to robust unmixing accounts for smooth temporal variations – construed as spectral variability – and abrupt spectral changes interpreted as outliers. The parameters of this model are inferred using an MCMC method, which allows the posterior of interest to be sampled and Bayesian estimators to be approximated. The encouraging results obtained on real data lead to the conclusion that MTHS images can provide improved unmixing performance when compared to methods analyzing the images independently.

Chapter 3. An online unmixing algorithm is proposed to analyze MTHS images, whose size may preclude the use of batch unmixing procedures. Based on a modified version of the model introduced in Chapter 2, the problem is formulated as a two-stage stochastic program, which is solved using a stochastic approximation. The results obtained on real data show the interest of considering several HS images to improve the unmixing results.

Chapter 4. The main practical contribution of this chapter results from the application of a recent asynchronous distributed algorithm to hyperspectral unmixing. The gain in terms of computational time brought by the asynchronous algorithm considered, when compared to its synchronous counterpart, is appreciable in the experiments conducted on synthetic and real data.

List of publications

▷ Submitted papers

[TDT17a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “A hierarchical Bayesian model accounting for endmember variability and abrupt spectral changes to unmix multitemporal hyperspectral images”. In: (May 2017). submitted. URL: http://thouvenin.perso.enseeiht.fr/papers/Thouvenin2017.pdf (cit. on p. 43).

[TDT17b] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Partially asynchronous distributed unmixing of hyperspectral images”. In: (2017). URL: https://arxiv.org/abs/1710.02574 (cit. on p. 99).

▶ International journals

[TDT16a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Hyperspectral unmixing with spectral variability using a perturbed linear mixing model”. In: *IEEE Trans. Signal Process.* 64.2 (Jan. 2016), pp. 525–538 (cit. on p. 17).

[TDT16b] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Online unmixing of multitemporal hyperspectral images accounting for spectral variability”. In: *IEEE Trans. Image Process.* 25.9 (Sept. 2016), pp. 3979–3990 (cit. on p. 75).

▶ International conferences

[TDT15a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “A perturbed linear mixing model accounting for spectral variability”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Nice, France, Sept. 2015, pp. 819–823.

[TDT16c] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Unmixing multitemporal hyperspectral images with variability: an online algorithm”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Shanghai, China, Mar. 2016, pp. 3351–3355.

[TDT17d] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Unmixing Multitemporal Hyperspectral Images Accounting for Smooth and Abrupt Variations”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Kos, Greece, Sept. 2017, pp. 2442–2446.

▶ National conferences

[TDT15b] 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*. in French. Lyon, France, Sept. 2015.

[TDT17c] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Une approche distribuée asynchrone pour la factorisation en matrices non-négatives – application au démélange hyperspectral”. In: *Actes du XXVIième Colloque GRETSI*. in French. Juan-les-Pins, France, Sept. 2017.

CHAPTER 1

Hyperspectral unmixing with spectral variability using a perturbed linear mixing model

This chapter has been adapted from the journal paper [TDT16a].

Contents

1.1. Introduction (in French)

Ce premier chapitre introduit un modèle paramétrique visant à prendre en compte la variabilité spatiale des spectres purs. Dans la mesure où les signatures qui composent les acquisitions peuvent être intréprétées comme autant de réalisations de spectres de références (caractéristiques de matériaux distincts), nous proposons une extension du modèle de mélange linéaire communément adopté. Tandis que la variabilité est exclusivement traduite en termes de facteurs d'échelle spatialement variables dans [Veg+14], le modèle proposé s'appuie sur une formulation inspirée du problème de moindres carrés totaux [GL80] et d'un modèle proposé dans un autre contexte [JJ13]. Chacune des signatures composant un pixel donné est représentée par un spectre pur, uniforme sur l'ensemble de l'image, affecté par une perturbation additive spatialement variable pour rendre compte de sa variabilité.

Les résultats prometteurs obtenus ces dernières années par la méthode des multiplicateurs de Lagrange (ADMM), aussi bien en imagerie hyperspectrale [Bio09] qu'en analyse ou débruitage d'images [ABF10; ABF11; AF13a; AF13b; MRF13], nous a dans un premier temps conduit à envisager une technique d'estimation basée sur cette méthode. De fait, nous considérons un algorithme de descente par coordonnée (BCD), dont les différents sous-problèmes fortement convexes sont résolus par la méthode ADMM. Par la suite, les performances de l'algorithme de minimisation proximale décrit dans [BST13; CPR16] sont évaluées dans le cadre du problème d'intérêt.

Le présent chapitre est organisé comme suit. Le modèle de mélange proposé est introduit dans un premier temps dans la section 1.3. La section 1.4 est consacrée à la description des deux algorithmes précédemment évoqués pour estimer les paramètres du modèle considéré. Les résultats de simulations sur données synthétiques et réelles sont respectivement rapportés dans les sections 1.5 et 1.6. Les résultats obtenus y sont systématiquement comparés à ceux d'autres algorithmes issus de l'état de l'art, comprenant deux méthodes de démélange linéaire (VCA/FCLS [NB05b; BF10] et SISAL/FCLS [Bio09]), et trois méthodes prenant en compte la variabilité spatiale des sources (AEB [Som+12], FDNS [JWZ10], et ELMM [Dru+16]). La section 1.7 permet enfin de conclure l'étude menée dans ce chapitre.

1.2. Introduction

In this chapter, an explicit model is introduced to address the question of spatial endmember variability. Since the identified endmembers can be construed as different instances of reference endmembers, we introduce an extended version of the classical LMM to explicitly model the observed variability. In [Veg+14], the variability is assumed to exclusively result from scaling factors. Resorting to an alternative interpretation of the problem, the proposed model is reminiscent of the total least squares (TLS) problem [GL80], and inspired by a model designed in [JJ13]. Each endmember is represented by a "pure" spectral signature, corrupted by an additive perturbation accounting for its variability. In the following, this model will be referred to as a perturbed linear mixing model (PLMM).

The promising results obtained with the alternating direction method of multipliers (ADMM) in hyperspectral imagery [Bio09] and in image deblurring [ABF10; ABF11; AF13a; AF13b; MRF13] serve as an incentive to apply a similar framework to estimate the parameters of the proposed model. To this end, a block coordinate descent (BCD) algorithm is first considered, whose strictly convex sub-problems are solved by ADMM. The performance of the proximal alternating linearized minimization (PALM) [BST13; CPR16] is also investigated for this problem.

This chapter is organized as follows. The PLMM accounting for spatial endmember variability is introduced in section 1.3. Section 1.4 first describes an ADMM-based algorithm to solve the resulting optimization problem, and then investigates the application of a PALM algorithm to the problem under study. Experimental results obtained on synthetic and real data are reported in section 1.5 and 1.6 respectively. The results obtained with the proposed algorithm are systematically compared to those of the vertex component analysis / fully constrained least squares (VCA/FCLS) [NB05b; BF10], the simplex identification via split augmented Lagrangian (SISAL) [Bio09] coupled with FCLS, the automated endmember bundles (AEB) [Som+12], the Fisher discriminant null space (FDNS) from [JWZ10], and the extended linear model (ELMM) introduced in [Dru+16]. Section 1.7 provides preliminary conclusions on this study.

1.3. Problem statement

1.3.1. A perturbed linear mixing model (PLMM)

In absence of specific prior knowledge on the variability nature, we propose to explicitly model endmember variability as an additive, spatially varying endmember perturbation. Reminiscent of the total least squares problem [GL80], this choice combines simplicity with a sufficient flexibility to account for the observed phenomenon. Assuming the number of endmembers R is known, the proposed PLMM differs from the classical LMM insofar as each pixel $\mathbf{y}_n$ is represented by a linear combination of the R endmembers – denoted by $\mathbf{m}_r$ – affected by a perturbation vector $\mathbf{dm}_{r,n}$ accounting for endmember variability. The resulting PLMM can be written

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\big(\mathbf{m}_r + \mathbf{dm}_{r,n}\big) + \mathbf{b}_n \text{ for } n = 1, \dots, N \tag{1.1}$$

where $\mathbf{y}_n$ denotes the nth image pixel, $\mathbf{m}_r$ is the rth endmember, $a_{r,n}$ is the proportion of the rth endmember in the nth pixel, and $\mathbf{dm}_{r,n}$ denotes the perturbation of the rth endmember in the nth pixel. Finally, $\mathbf{b}_n$ is a zero-mean white Gaussian noise resulting from the data acquisition and modeling errors. In absence of variability, the PLMM reduces to the LMM. The PLMM (1.1) can be written in matrix form

$$\mathbf{Y} = \mathbf{MA} + \underbrace{\big[\mathbf{dM}_1\mathbf{a}_1, \dots, \mathbf{dM}_N\mathbf{a}_N\big]}_{\mathbf{\Delta}} + \mathbf{B} \tag{1.2}$$

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

$$\begin{aligned} &\mathbf{A} \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}^{\mathrm{T}}\mathbf{1}_R = \mathbf{1}_N \\ &\mathbf{M} \succeq \mathbf{0}_{L,R}, \quad \mathbf{M} + \mathbf{dM}_n \succeq \mathbf{0}_{L,R}, \ \forall n = 1, \dots, N \end{aligned} \tag{1.3}$$

where $\succeq$ denotes a term-wise inequality. When compared to other approaches previously proposed in the literature [ZH14], the model (1.1) presents the advantage of explicitly addressing spectral variability in terms of an additive perturbation. As illustrated in the experiments reported in section 1.6, this perturbation accounts for any deviation from the linear mixing model. In the following paragraphs, we propose an unsupervised algorithm to estimate the endmembers, the abundances and the endmember variability for each pixel composing the image of interest using the model (1.1) with the constraints (1.3).

Using an alternative approach, another recently proposed model explicitly represents endmember variability in terms of spatially varying scaling factors [Dru+16]. Note that the model used in practice in [Dru+16] can be reformulated as a linear unmixing problem without the abundance sum-to-one constraint (more precisely, the sum of the abundance coefficient of each pixel is equal to the corresponding unknown scale coefficient, see (I.7) in the preliminary chapter). In practice, the resulting scale indeterminacy is partly addressed by considering an additional term, which penalizes the distance between reference endmembers and the signatures estimated for each pixel (see (I.8)).

Remark 1.1. The proposed model exhibits similarities with the NCM introduced in [Ech+10] and its generalization, the GNCM, proposed in [HDT15]. The GNCM is expressed as

$$\mathbf{y}_n = \sum_r a_{r,n}\mathbf{m}_{r,n} + \mathbf{e}_n, \quad \mathbf{e}_n \sim \mathcal{N}(\mathbf{0}, \psi_n^2\mathbf{I}), \quad \mathbf{s}_{r,n} \sim \mathcal{N}(\mathbf{m}_r, \operatorname{diag}(\boldsymbol{\sigma}^2)) \quad (\boldsymbol{\sigma}^2 \in \mathbb{R}_+^L).$$

The PLMM and the GNCM are equivalent by setting $\psi_n = 0$, and considering $\mathbf{m}_{r,n} = \mathbf{m}_r + \mathbf{dm}_{r,n}$ with $\mathbf{dm}_{r,n} \sim \mathcal{N}(\mathbf{0}, \operatorname{diag}(\boldsymbol{\sigma}^2))$. The distinction between the work proposed in this chapter and [HDT15] lies in the estimation approach developed in the following section, and in the nature of the parameter estimators. More precisely, the approach proposed in [HDT15] directly considers the likelihood function

$$\begin{aligned} &p(\mathbf{y}_n \mid \mathbf{A}, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) \propto \Big(\prod_{\ell=1}^{L} \lambda_{\ell,n}\Big)^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}\big\langle \operatorname{diag}(\boldsymbol{\lambda}_n)\big(\mathbf{y}_n - \mathbf{M}\mathbf{a}_n\big), \mathbf{y}_n - \mathbf{M}\mathbf{a}_n\big\rangle\Big\} \\ &\lambda_{\ell,n} = \Big(\sum_{r=1}^{R} a_{r,n}^2\sigma_{r,\ell}^2 + \psi_n^2\Big)^{-1}, \quad \boldsymbol{\lambda}_n = \big[\lambda_{\ell,n}\big]_\ell \in \mathbb{R}^L \end{aligned}$$

where the variability does not explicitly appear (since contained in $\mathbf{m}_{r,n}$). On the contrary, we propose to explicitly estimate the variability present in the scene to characterize its spatial distribution within the scene under study.

1.3.2. Problem formulation

The PLMM (1.1) and the constraints (1.3) can be combined to formulate a constrained optimization problem expressed as

$$(\mathbf{M}^*, \mathbf{dM}^*, \mathbf{A}^*) \in \operatorname*{arg\,min}_{\mathbf{M}, \mathbf{dM}, \mathbf{A}} \Big\{\mathcal{J}(\mathbf{M}, \mathbf{dM}, \mathbf{A}) \text{ s.t. } (1.3)\Big\}. \tag{1.4}$$

Assuming the signal is corrupted by a zero-mean white Gaussian noise, the objective function can be defined as

$$\mathcal{J}(\mathbf{M}, \mathbf{dM}, \mathbf{A}) = F(\mathbf{M}, \mathbf{dMA}) + \alpha\Phi(\mathbf{A}) + \beta\Psi(\mathbf{M}) + \gamma\Upsilon(\mathbf{dM}) \tag{1.5}$$

$$F(\mathbf{M}, \mathbf{dM}, \mathbf{A}) = \sum_n f(\mathbf{M}, \mathbf{dM}_n, \mathbf{a}_n) = \frac{1}{2}\left\|\mathbf{Y} - \mathbf{MA} - \boldsymbol{\Delta}\right\|_{\mathrm{F}}^2, \quad \boldsymbol{\Delta} = \big[\mathbf{dM}_1\mathbf{a}_1, \dots, \mathbf{dM}_N\mathbf{a}_N\big] \tag{1.6}$$

where the expression of the data fitting term F, given by the Frobenius distance between the acquisitions $\mathbf{Y}$ and the reconstructed data $\mathbf{MA} + \boldsymbol{\Delta}$, results from the assumption on the noise statistical distribution. The penalization functions Φ, Ψ and Υ reflect any *a priori* knowledge related to the mixture parameters $\mathbf{M}$, $\mathbf{A}$ and $\mathbf{dM}$, whereas the hyper-parameters α, β, γ control the trade-off between the data fitting and the penalization terms. We further assume that the penalization functions are separable as follows

$$\Phi(\mathbf{A}) = \sum_{n=1}^{N} \phi(\mathbf{a}_n), \quad \Psi(\mathbf{M}) = \sum_{\ell=1}^{L} \psi(\widetilde{\mathbf{m}}_\ell), \quad \Upsilon(\mathbf{dM}) = \sum_{n=1}^{N} \upsilon(\mathbf{dM}_n) \tag{1.7}$$

where $\widetilde{\mathbf{m}}_\ell$ denotes the ℓth row of $\mathbf{M}$, and ϕ, ψ and υ are non-negative differentiable convex functions. In the following, this assumption is notably used to decompose (1.4) into a collection of simpler sub-problems, further described in section 1.4.1. The penalization functions considered in this chapter are introduced in the next paragraphs.

Abundance penalization. To promote spatially smooth abundances as in [CRH14], we consider a penalization function Φ expressed as

$$\Phi(\mathbf{A}) = \frac{1}{2} \|\mathbf{AH}\|_{\mathrm{F}}^2 \tag{1.8}$$

where $\mathbf{H}$ is an operator computing the differences between the abundances of a given pixel and those of its 4 nearest neighbors. The resulting expression of ϕ is detailed in Appendix A.1.

Endmember penalization. The choice of Ψ is based on classical penalization functions found in the literature, which essentially consist in constraining the volume of the simplex whose vertices are the endmembers. Notably, the penalization used in [MQ07; Cha+11] enables the volume exactly occupied by the $(R-1)$-simplex formed by the endmembers to be penalized. Another approach consists in approximating this volume by the mutual distance between the endmembers [Ber+04; ASL11]. Finally, if the endmembers are *a priori* expected to be close to reference spectral signatures, a penalization on the distance between the endmembers and these signatures can be considered. In the following lines, the expression of these three penalization terms is recalled, and the expression of the corresponding functions ψ is given in Appendix A.1.

1. The distance between the endmembers and reference spectral signatures $\mathbf{M}_0$ can be expressed as

$$\Psi(\mathbf{M}) = \frac{1}{2} \|\mathbf{M} - \mathbf{M}_0\|_{\mathrm{F}}^2 . \tag{1.9}$$

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

$$\Psi(\mathbf{M}) = \frac{1}{2} \sum_{i=1}^{R} \left(\sum_{\substack{j=1 \\ j \neq i}}^{R} \|\mathbf{m}_i - \mathbf{m}_j\|_2^2 \right). \tag{1.10}$$

3. Under the pure pixel and linear mixture assumptions, the data points are enclosed in an $(R-1)$-simplex whose vertices are the endmembers [Cha+11]. Let $\mathbf{E}$ be the projection of $\mathbf{M}$ on the subspace spanned by the $R-1$

principal components of $\mathbf{Y}$. The expression of the volume of this subspace is

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

To ensure the differentiability of the penalization term with respect to $\mathbf{E}$, we propose to consider

$$\Psi(\mathbf{M}) = \frac{1}{2}\mathcal{V}^2(\mathbf{E}). \tag{1.11}$$

Variability penalization. Finally, the variability penalizing function Υ has been chosen to limit the norm of the captured variability. Indeed, endmember variability is expected to represent a limited fraction of the signal's energy. The following penalty function, which has the advantage to be differentiable with respect to $\mathbf{dM}_n$, has thus been selected

$$\Upsilon\left(\mathbf{dM}\right) = \frac{1}{2}\sum_{n=1}^{N}\left\|\mathbf{dM}_n\right\|_{\mathrm{F}}^2. \tag{1.12}$$

To the best of our knowledge, no specific information regarding the spatial distribution of the variability is available in the remote sensing literature. Consequently, no specific regularization has been considered on $\mathbf{dM}$. However, any spatial penalization satisfying the assumptions given in paragraph 1.3.2 can be considered if necessary (e.g., a group-Lasso $\ell_{2,1}$ penalization to promote spatial sparsity of the variability term $\mathbf{dM}$).

Note that an alternative choice of Υ has been considered in section 1.4.2 by defining $\upsilon(\mathbf{dM}_n) = \iota_{\{\|\cdot\|_{\mathrm{F}}\leq\nu\}}(\mathbf{dM}_n)$, where $\iota_{\mathcal{S}}$ denotes the indicator function of a given set $\mathcal{S}$ and $\nu > 0$ is a user-defined constant. In practice, the choice of ν results from a compromise between the amount a variability captured by the model, and the risk to capture noise into the unmixing process.

1.4. Parameter estimation

The parameters involved in the model (1.1) can be estimated by different algorithms. Paragraph 1.4.1 is focused on a *block coordinate descent* (BCD) algorithm, leveraging the solutions to simpler strongly convex subproblems obtained by the ADMM. An application of the PALM algorithm is then investigated in paragraph 1.4.2.

1.4.1. An ADMM-based algorithm

Given the non-convexity of the problem (1.4), an alternating minimization strategy similar to [AF13a] 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 Φ, Ψ, Υ in section 1.3.1 allow the optimization problems described in Algo. 1 to be divided into a collection of strictly convex sub-problems. These sub-problems involve differentiable functions, thus simplifying their resolution. Having introduced appropriate splitting variables to account for

Algorithm 1: PLMM-unmixing: a BCD/ADMM algorithm. Each sub-problem resulting from the decomposition of the optimization steps (a), (b), and (c) is solved by ADMM.

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

the constraints, these sub-problems are finally solved using ADMM steps, which admit 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 A.3 for the sake of brevity.

Abundance estimation (A). With the assumptions made in paragraph 1.3.2, optimizing the cost function $\mathcal{J}$ with respect to $\mathbf{A}$ under the constraints (1.3) is equivalent to solving the following problems

$$\mathbf{a}_n^* = \arg\min_{\mathbf{a}_n} \left\{ \begin{array}{c} \frac{1}{2}\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\|_2^2 + \alpha\phi(\mathbf{a}_n) \\ \\ \text{s.t.} \quad \mathbf{a}_n \succeq \mathbf{0}_R, \quad \mathbf{a}_n^{\mathrm{T}}\mathbf{1}_R = 1 \end{array} \right\}. \tag{1.13}$$

Introducing the splitting variables $\mathbf{w}_n^{(\mathbf{A})} \in \mathbb{R}^R$ for $n = 1, \dots, N$ such that $\mathbf{a}_n = \mathbf{w}_n^{(\mathbf{A})}$, the resulting scaled augmented Lagrangian is expressed as

$$\begin{aligned} \mathcal{L}_{\mu_n^{(\mathbf{A})}}\left(\mathbf{a}_n, \mathbf{w}_n^{(\mathbf{A})}, \boldsymbol{\lambda}_n^{(\mathbf{A})}, \lambda_n\right) = \frac{1}{2}\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\|_2^2 + \frac{\mu_n^{(\mathbf{A})}}{2}\|\mathbf{a}_n - \mathbf{w}_n^{(\mathbf{A})} + \boldsymbol{\lambda}_n^{(\mathbf{A})}\|_2^2 \\ + \lambda_n(\mathbf{1}_R^{\mathrm{T}}\mathbf{a}_n - 1) + \alpha\phi(\mathbf{a}_n) + \iota_{\{\cdot \succeq \mathbf{0}_R\}}\left(\mathbf{w}_n^{(\mathbf{A})}\right) \end{aligned} \tag{1.14}$$

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 A.3.

Endmember estimation (M). Similarly to the previous paragraph, optimizing $\mathcal{J}$ with respect to $\mathbf{M}$ under the constraint (1.3) is equivalent to solving

$$\widetilde{\mathbf{m}}_\ell^* = \arg\min_{\widetilde{\mathbf{m}}_\ell} \left\{ \frac{1}{2}\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\|_2^2 + \beta\psi(\widetilde{\mathbf{m}}_\ell), \text{ s.t. } \widetilde{\mathbf{m}}_\ell \succeq \widetilde{\mathbf{c}}_\ell \right\} \tag{1.15}$$

Algorithm 2: ADMM optimization w.r.t. $\mathbf{A}$ (step ((a)) of Algo. 1).

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

where $\widetilde{\mathbf{m}}_\ell$ denotes the ℓth row of $\mathbf{M}$, $\widetilde{\mathbf{c}}_\ell = \max\{\mathbf{0}_L^{\mathrm{T}}, \max_n\{-\widetilde{\mathbf{dm}_{\ell,n}}\}\}$ and the max operations are applied term-wise. Introducing the splitting variables $\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})} \in \mathbb{R}^{1\times L}$ such that $\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})} = \widetilde{\mathbf{m}}_\ell$ for $\ell = 1, \ldots, L$ leads to consider the following scaled augmented Lagrangian

$$\begin{aligned}\mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}, \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right) = \frac{1}{2}\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A} - \widetilde{\boldsymbol{\delta}}_\ell\|_2^2 + \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})} + \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right\|_2^2 \\ + \beta\psi(\widetilde{\mathbf{m}}_\ell) + \iota_{\{\cdot\succeq\widetilde{\mathbf{c}}_\ell\}}\left(\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}\right)\end{aligned} \tag{1.16}$$

with $\mu_\ell^{(\mathbf{M})} > 0$. The resulting algorithm (step (b) of Algo. 1) is similar to Algo. 2. The solution to these optimization problems, which depends on the selected penalizing function Ψ introduced in paragraph 1.3.2, is reported in Appendix A.3.

Variability estimation (dM). Finally, optimizing $\mathcal{J}$ with respect to $\mathbf{dM}$ under the constraint (1.3) is equivalent to solving the sub-problems

$$\mathbf{dM}_n^* = \arg\min_{\mathbf{dM}_n}\left\{\frac{1}{2}\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\|_2^2 + \gamma\upsilon(\mathbf{dM}_n), \text{ s.t. } \mathbf{M} + \mathbf{dM}_n \succeq \mathbf{0}_{L,R}\right\}. \tag{1.17}$$

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

$$\begin{aligned}\mathcal{L}_{\mu_n^{(\mathbf{dM})}}\left(\mathbf{dM}_n, \mathbf{W}_n^{(\mathbf{dM})}, \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right) = \frac{1}{2}\|\mathbf{y}_n - (\mathbf{M} + \mathbf{dM}_n)\mathbf{a}_n\|_2^2 + \gamma\upsilon(\mathbf{dM}_n) + \iota_{\{\cdot\succeq-\mathbf{M}\}}\left(\mathbf{W}_n^{(\mathbf{dM})}\right) \\ + \frac{\mu_n^{(\mathbf{dM})}}{2}\left\|\mathbf{dM}_n - \mathbf{W}_n^{(\mathbf{dM})} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right\|_{\mathrm{F}}^2\end{aligned} \tag{1.18}$$

Algorithm 3: PALM algorithm to solve the unmixing problem (1.4).

```
Data: $\mathbf{Y}$, $\mathbf{A}^{(0)}$, $\mathbf{M}^{(0)}$, $\mathbf{dM}^{(0)}$
begin
    $k \leftarrow 0$;
    while stopping criterion not satisfied do
        // Abundance update
        for $n = 1$ to $N$ do
(a)         $\mathbf{a}_n^{k+1} = \text{prox}_{\iota_{\mathcal{S}_R}}\left(\mathbf{a}_n^k - \frac{1}{\eta_n^k}\nabla_{\mathbf{a}_n} f\left(\mathbf{a}_n^k, \mathbf{M}^k, \mathbf{dM}_n^k\right)\right)$;    // (1.19)
        // Endmember update
(b)     $\mathbf{M}^{k+1} = \text{prox}_{\iota_{\{\cdot \succeq \mathbf{C}^k\}}}\left(\mathbf{M}^k - \frac{1}{\mu^k}\nabla_{\mathbf{M}} F\left(\mathbf{A}^{k+1}, \mathbf{M}^k, \mathbf{dM}^k\right)\right)$;    // (1.21)
        // Variability update
        for $n = 1$ to $N$ do
(c)         $\mathbf{dM}_n^{k+1} = \text{prox}_{\iota_{\{\|\cdot\|_F \leq \nu\}} + \iota_{\{\cdot \succeq -\mathbf{M}^{k+1}\}}}\left(\mathbf{dM}_n^k - \frac{1}{\nu_n^k}\nabla_{\mathbf{dM}_n} f\left(\mathbf{a}_n^{k+1}, \mathbf{M}^{k+1}, \mathbf{dM}_n^k\right)\right)$;    // (1.23)
        $k \leftarrow k + 1$;
Result: $\mathbf{A}^k$, $\mathbf{M}^k$, $\mathbf{dM}^k$
```

with $\mu_n^{(\mathbf{dM})} > 0$. The resulting algorithm (step (c) of Algo. 1) is similar to Algo. 2. The solution to these sub-problems is given in Appendix A.3.

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.

Convergence and computational cost. The BCD proposed in Alg. 1 is guaranteed to converge to a stationary point of the objective function $\mathcal{J}$ as long as each sub-problem is exactly minimized [Ber99, Proposition 2.7.1]. Besides, the sub-problems tackled in the preceding paragraphs are strongly convex, ensuring 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 [Boy+10]). The same convergence result applies to the ADMM with the parameter adjustment reported in Appendix A.2 as long as the parameter is updated finitely many times [Boy+10]. Considering the significant number of unknown parameters and the simple expression of the ADMM updates detailed in Appendix A.3, we can note that the computational cost is dominated by matrix products, yielding an overall computational cost of the order $\mathcal{O}(LR^2N)$.

We may however mention that the proximal alternating linearized minimization (PALM) [BST13] can also be directly applied to the considered problem, thus allowing a rigorous convergence result based on the Kurdyka-Łojasiewicz property to be obtained. This alternative approach is investigated in the following section.

1.4.2. Parameter estimation based on the PALM algorithm

A direct application of [BST13, Remark 4 (iv)] leads to the proximal gradient descent detailed in Algo. 3. The following lines give further details on the expressions used to update the different parameters of interest.

Abundance update (A). Updating the abundance coefficients in the absence of any penalization reduces to the following proximal gradient step

$$\mathbf{a}_n^{k+1} = \operatorname{prox}_{\iota_{\mathcal{S}_R}}\left(\mathbf{a}_n^k - \frac{1}{\eta_n^k}\nabla_{\mathbf{a}_n} f\left(\mathbf{a}_n^k, \mathbf{M}^k, \mathbf{dM}_n^k\right)\right), \quad \eta_n^k = \left\|(\mathbf{M}^k + \mathbf{dM}_n^k)^{\mathrm{T}}(\mathbf{M}^k + \mathbf{dM}_n^k)\right\|_{\mathrm{F}}. \tag{1.19}$$

When considering the spatial smoothness regularization introduced on section 1.3.2, the update step becomes

$$\begin{aligned} \mathbf{A}^{k+1} &= \operatorname{prox}_{\iota_{\mathcal{A}_R}}\left(\mathbf{A}^k - \frac{1}{\eta^k}\nabla_{\mathbf{A}} F\left(\mathbf{A}^k, \mathbf{M}^k, \mathbf{dM}^k\right)\right) \\ \eta^k &= \max_n\left\{\left\|(\mathbf{M}^k + \mathbf{dM}_n^k)^{\mathrm{T}}(\mathbf{M}^k + \mathbf{dM}_n^k)\right\|_2\right\} + \alpha\left\|\mathbf{H}\mathbf{H}^{\mathrm{T}}\right\|_{\mathrm{F}}. \end{aligned} \tag{1.20}$$

Indeed, a Lipschitz constant can be explicitly computed for $\nabla_{\mathbf{A}} F\left(\mathbf{A}^k, \mathbf{M}^k, \mathbf{dM}^k\right)$ by observing that, for two given matrices $\mathbf{A}_1$, $\mathbf{A}_2$,

$$\begin{aligned} \left\|\left[\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1(\mathbf{a}_1^1 - \mathbf{a}_1^2), \dots, \mathbf{M}_N^{\mathrm{T}}\mathbf{M}_N(\mathbf{a}_N^1 - \mathbf{a}_N^2)\right]\right\|_{\mathrm{F}}^2 &= \sum_n \left\|\mathbf{M}_n^{\mathrm{T}}\mathbf{M}_n(\mathbf{a}_n^1 - \mathbf{a}_n^2)\right\|_2^2 \\ &\leq \sum_n \left\|\mathbf{M}_n^{\mathrm{T}}\mathbf{M}_n\right\|_2^2 \left\|\mathbf{a}_n^1 - \mathbf{a}_n^2\right\|_2^2 \\ &\leq \left(\max_n \left\|\mathbf{M}_n^{\mathrm{T}}\mathbf{M}_n\right\|_2^2\right)\left\|\mathbf{A}_1 - \mathbf{A}_2\right\|_{\mathrm{F}}^2. \end{aligned}$$

Endmember update (M). Considering the constraints (1.3), the endmember update reduces to the following projected gradient step

$$\begin{aligned} \mathbf{M}^{k+1} &= \operatorname{prox}_{\iota_{\{\cdot \succeq \mathbf{C}^k\}}}\left(\mathbf{M}^k - \frac{1}{\mu^k}\nabla_{\mathbf{M}} F\left(\mathbf{A}^{k+1}, \mathbf{M}^k, \mathbf{dM}^k\right)\right) \\ \mathbf{C}^k &= \max\left\{\mathbf{0}_{L,R}, \max_n -\mathbf{dM}_n^k\right\} \end{aligned} \tag{1.21}$$

where the expression of μ^k depends on the penalization function considered

$$\mu^k = \begin{cases} \left\|\mathbf{A}^{k+1}(\mathbf{A}^{k+1})^{\mathrm{T}}\right\|_{\mathrm{F}} & \text{(no penalization)} \\ \left\|\mathbf{A}^{k+1}(\mathbf{A}^{k+1})^{\mathrm{T}} + \beta\mathbf{G}\mathbf{G}^{\mathrm{T}}\right\|_{\mathrm{F}} & \text{(mutual distance)} \\ \left\|\mathbf{A}^{k+1}(\mathbf{A}^{k+1})^{\mathrm{T}} + \beta\mathbf{I}_R\right\|_{\mathrm{F}} & \text{(distance to a reference } \mathbf{M}_0\text{).} \end{cases} \tag{1.22}$$

Variability update (dM). Similarly, the variability update step can be written

$$\mathbf{dM}_n^{k+1} = \operatorname{prox}_{\iota_{\{\|\cdot\|_{\mathrm{F}} \leq \nu\}} + \iota_{\{\cdot \succeq -\mathbf{M}^{k+1}\}}}\left(\mathbf{dM}_n^k - \frac{1}{\nu_n^k}\nabla_{\mathbf{dM}_n} f\left(\mathbf{a}_n^{k+1}, \mathbf{M}^{k+1}, \mathbf{dM}_n^k\right)\right), \quad \nu_n^k = \left\|\mathbf{a}_n^{k+1}(\mathbf{a}_n^{k+1})^{\mathrm{T}}\right\|_{\mathrm{F}}. \tag{1.23}$$

The above proximal operator can be easily computed as (see Appendix A.4)

$$\operatorname{prox}_{\iota_{\{\|\cdot\|_{\mathrm{F}}\leq\nu\}}+\iota_{\{\cdot\succeq-\mathbf{M}^{k+1}\}}}(\mathbf{X}) = (\operatorname{prox}_{\iota_{\{\|\cdot\|_{\mathrm{F}}\leq\nu\}}} \circ \operatorname{prox}_{\iota_{\{\cdot\succeq-\mathbf{M}^{k+1}\}}})(\mathbf{X}). \tag{1.24}$$

Convergence and computational cost. The problem (1.4) satisfies the conditions of [BST13, Theorem 1] for the objective function (1.5) (e.g., as a special case of [BST13, Section 4.1]). Consequently, the sequence of iterates $\{\mathbf{M}^k, \mathbf{dM}^k, \mathbf{A}^k\}_k$ is guaranteed to converge to a critical point of the unmixing problem (1.4). Dominated by matrix products, the per iteration complexity of the PALM algorithm is of the order of $\mathcal{O}(LRN)$, since $R \ll L \ll N$ in practice.

1.5. Experiments with synthetic data

The performance of the proposed method is assessed on four images of size 128×64, composed of 413 bands. Each image corresponds to a mixture of R endmembers in presence or absence of pure pixels, with $R \in \{3, 6\}$. The data have been further corrupted by an additive white Gaussian noise to ensure a signal-to-noise ratio (SNR) of $30\,\mathrm{dB}$.

Since no commonly accepted variability model is available in the literature, we propose the following procedure to generate controlled variability. The perturbed endmembers involved in the mixture have been generated using the component-wise product of reference endmembers with randomly drawn piece-wise affine functions, providing realistic perturbed endmembers as illustrated in Fig. 1.1. This choice allows the shape and amplitude of the reference endmembers to be easily modified in different spectral bands, while generating satisfactory perturbed endmembers when compared to the variability observed in real data (see Fig. I.2(a) in the preliminary chapter). For simplicity, we consider piece-wise affine functions composed of two parts delimited by the band index L_{break}, where a parameter c_{var} controls the amplitude of the extreme points composing each of the two functions (see Fig. 1.2).

For a given variability coefficient $c_{\text{var}} > 0$ fixed by the user, the parameters ξ_i, $i \in \{1, 2, 3\}$ and $L_{\text{break}} \in \{1, \dots, L\}$ introduced in Fig. 1.2 have been independently drawn as follows

$$\xi_i \sim \mathcal{U}_{[1-c_{\text{var}}/2, 1+c_{\text{var}}/2]}, \; i \in \{1, 2, 3\} \tag{1.25}$$

$$L_{\text{break}} = \lfloor L/2 + \lfloor LU/3 \rfloor \rfloor, \; U \sim \mathcal{N}(0, 1) \tag{1.26}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The synthetic data used in the proposed experiments have been generated with a value of c_{var} that is lower in the upper half of the image ($c_{\text{var}} = 0.1$) than in the lower half ($c_{\text{var}} = 0.25$). Note that different affine functions have been considered in each pixel for each endmember.

1.5.1. State-of-the-art methods

The results of the proposed algorithms, evaluated in terms of the performance measures introduced in the preliminary chapter (see section I.4), have been compared to those obtained with two classical linear unmixing methods (VCA/FCLS, SISAL/FCLS) and three variability accounting algorithms (AEB, FDNS, ELMM). These methods are recalled below with

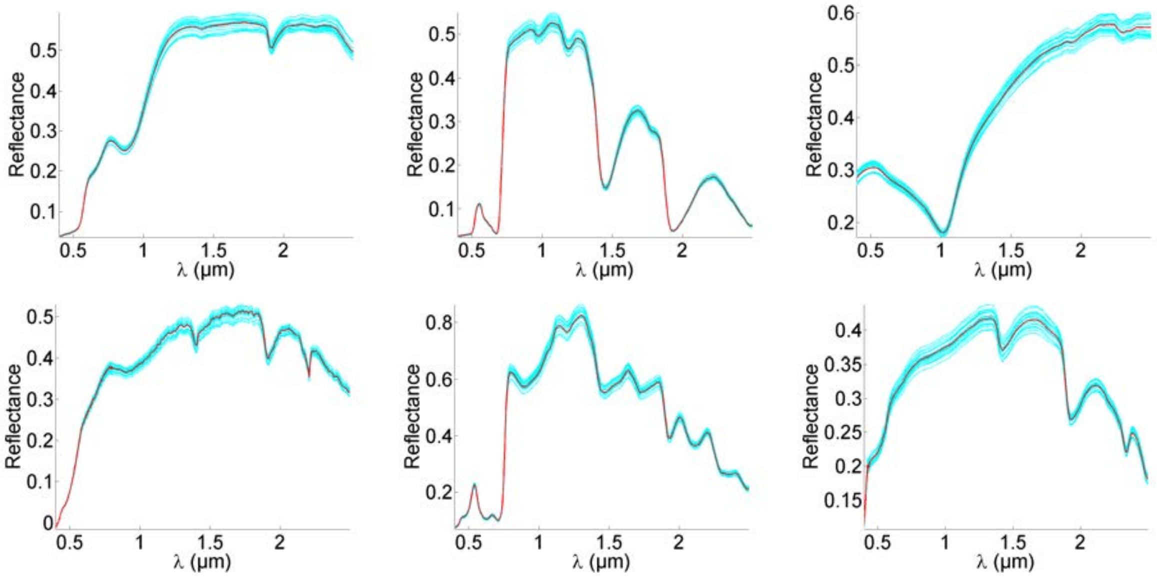


Figure 1.1.: Reference endmembers (red lines) and 20 corresponding instances under spectral variability (cyan lines) involved in the synthetic data experiments.

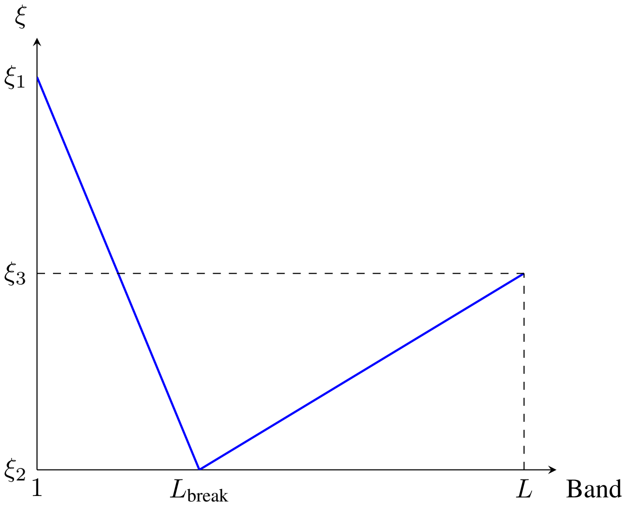


Figure 1.2.: Example of a randomly-generated affine function used to generate the synthetically perturbed endmembers.

their most relevant implementation details.

- Classical unmixing methods (no variability)
    1. VCA/FCLS: the endmembers are first extracted using the vertex component analysis [NB05b]. The abundances are then estimated for each pixel using the fully constrained least squares (FCLS) algorithm [HC00; BF10]);
    2. SISAL/FCLS: the endmembers are first extracted using the simplex identification via split augmented Lagrangian [Bio09]. The tolerance for the stopping rule has been set to 10^{-3} and VCA has been used as an initialization step. The abundances are then estimated for each pixel using FCLS.
- Variability accounting methods
    3. AEB [Som+12; Rob+98; Goe+13]: the size of the bundles is equal to 25% of the total pixel number. The endmembers and abundance are estimated using VCA/FCLS;
    4. FDNS [JWZ10]: the endmembers and abundances are first estimated by VCA/FCLS;
    5. ELMM [Dru+16]: initialized with VCA/FCLS, a slightly modified version of the BCD algorithm proposed in [Dru+16] has been applied (essentially to correct the update equation [Dru+16, (12)]). The algorithm is stopped when the relative difference between two successive values of the objective function is less than 10^{-4}, with a maximum of $N_{\text{iter}} = 100$ iterations. To allow a better comparison with the proposed methods, the variability captured by the ELMM is interpreted in Fig. 1.4 as the deviation of each spectral signature from its average;
    6. 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 if the imaged scene does not satisfy this assumption. The variability matrices have been initialized with all their entries equal to 0. The algorithm is stopped when the relative difference between two successive values of the objective function is less than 10^{-4}, with a maximum of $N_{\text{iter}} = 100$ iterations.
    7. Proposed method (PALM): the algorithm is stopped when the relative difference between two successive values of the objective function is less than 10^{-5}, with a maximum of $N_{\text{iter}} = 500$ iterations.

In the following, only the results obtained with a combination of the abundance spatial smoothness (abbreviated as *ss* in the following) and the endmember mutual distance (*md*) regularizations are reported for BCD/ADMM and PALM. Different regularization combinations (i.e., with the regularization based on the minimum volume or the distance to reference endmembers) were not found to bring any significant improvement over the reported results.

1.5.2. Results

The parameters used for the ADMM algorithms are detailed in Table 1.1, and the values chosen for α, β, γ and ν (for BCD/ADMM and/or PALM) are reported in Table 1.2 and 1.3. The values for these parameters have been chosen using

Table 1.1.: ADMM parameters.

	Synthetic data	Moffett scene	Cuprite scene
τ^{incr}	1.1	1.1	1.1
τ^{decr}	1.1	1.1	1.1
μ	10	10	10
$\mu_n^{(\mathbf{A})(0)}$	100	100	100
$\mu_\ell^{(\mathbf{M})(0)}$	10^{-1}	1	10^{-1}
$\mu_n^{(\mathbf{dM})(0)}$	10^{-1}	1	10^{-1}
ε^{abs}	10^{-2}	10^{-2}	10^{-2}
ε^{rel}	10^{-4}	10^{-4}	10^{-4}
N_{iter}	100	100	100

a grid of values successively tested for each parameter, while the others are fixed. The value selected for each parameter corresponds to the one minimizing the reconstruction error RE. The values taken to initialize this procedure ensure that each penalization term has a weight corresponding to a predefined fraction of the initial data fitting term (typically 10 to 20 %).

The performance measures returned by the unmixing methods, provided in Table 1.2 for the datasets containing pure pixels and in Table 1.3 for datasets without pure pixels, led 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 scenarii, the abundance GMSEs and the REs are better or comparable to those resulting from state-of-the-art methods. More specifically, the ELMM yields satisfactory performance resulting from the spatial prior considered in [Dru+16]. In addition, AEB and SISAL/FCLS lead to satisfactory performances in presence of pure pixels, which partly results from the underpinning use of VCA/FCLS (which is a pure-pixel based methods);
- The proposed method is computationally more expensive than existing algorithms,, which results in part from the number of additional parameters considered in the PLMM.

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 the VCA/FCLS algorithm is used as an initialization step.

Finally, the variability captured by ELMM and the proposed model is presented in Figs. 1.4 and 1.6 for one of the synthetic datasets (no pure pixels, $R = 3$). The difference between the variability intensities detected in the upper and the lower part of the scene is due to the different variability coefficients used to generate these areas. A similar observation can be made on the scaling factors estimated by ELMM, thus illustrating the consistency of the results obtained with the ELMM and the proposed approach.

Table 1.2.: Simulation results on synthetic data **in presence of pure pixels** (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$), $[(\alpha, \beta) = (1.35, 1.15 \times 10^{-5})$ for $R = 3$, $(\alpha, \beta) = (3.7 \times 10^{-1}, 4.9 \times 10^{-6})$ for $R = 6]$.

		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	time (s)
$R = 3$	VCA/FCLS	6.38	3.80	/	7.56	**1**
	SISAL/FCLS	5.27	3.08	/	3.35	2
	FDNS	6.00	3.79	/	7.56	4
	AEB	5.70	**2.07**	/	3.50	52
	ELMM	5.93	2.69	7.15	**0.58**	390
	ssmdBCD/ADMM ($\gamma = 1$)	**5.10**	2.56	4.46	1.43	729
	ssmdPALM ($\nu = 10^{-3}$)	5.30	3.51	**3.41**	3.63	307
$R = 6$	VCA/FCLS	6.33	2.24	/	2.92	**1**
	SISAL/FCLS	**3.84**	3.05	/	2.25	3
	FDNS	6.33	2.22	/	2.92	5
	AEB	5.70	1.31	/	2.40	142
	ELMM	6.26	**0.84**	4.18	**0.42**	501
	ssmdBCD/ADMM ($\gamma = 1$)	5.48	1.54	2.88	0.77	224
	ssmdPALM ($\nu = 7.2 \times 10^{-4}$)	4.84	1.99	**2.74**	1.90	457

Table 1.3.: Simulation results on synthetic data **in absence of pure pixels** (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$) $[(\alpha, \beta) = (2.1 \times 10^{-1}, 7.7 \times 10^{-6})$ for $R = 3$, $(\alpha, \beta) = (7.1 \times 10^{-1}, 4.3 \times 10^{-6})$ for $R = 6]$.

		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	time (s)
$R = 3$	VCA/FCLS	5.06	2.07	/	2.66	**1**
	SISAL/FCLS	4.43	2.16	/	2.56	2
	FDNS	5.06	2.06	/	2.66	3
	AEB	5.11	2.11	/	2.66	33
	ELMM	5.05	1.78	6.86	4.34	329
	ssmdBCD/ADMM ($\gamma = 10^{-1}$)	4.56	**1.49**	6.21	**0.08**	285
	ssmdPALM ($\nu = 5 \times 10^{-2}$)	**4.51**	1.54	**5.24**	0.60	314
$R = 6$	VCA/FCLS	6.55	2.52	/	2.82	**4**
	SISAL/FCLS	6.04	**1.63**	/	2.02	5
	FDNS	6.55	2.53	/	2.82	7
	AEB	**6.00**	1.78	/	1.85	208
	ELMM	6.54	1.98	4.13	**0.60**	555
	ssmdBCD/ADMM ($\gamma = 1$)	6.19	2.19	2.89	0.81	618
	ssmdPALM ($\nu = 2 \times 10^{-1}$)	6.05	2.21	**2.73**	1.82	449

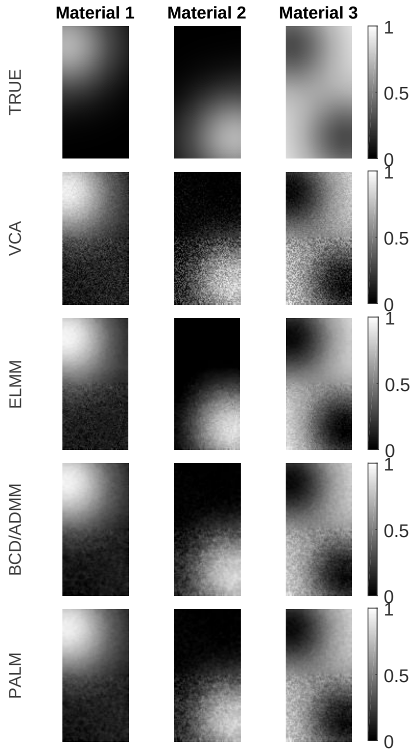


Figure 1.3.: Estimated abundance maps obtained from the synthetic dataset in absence of pure pixels, composed of $R = 3$ endmembers.

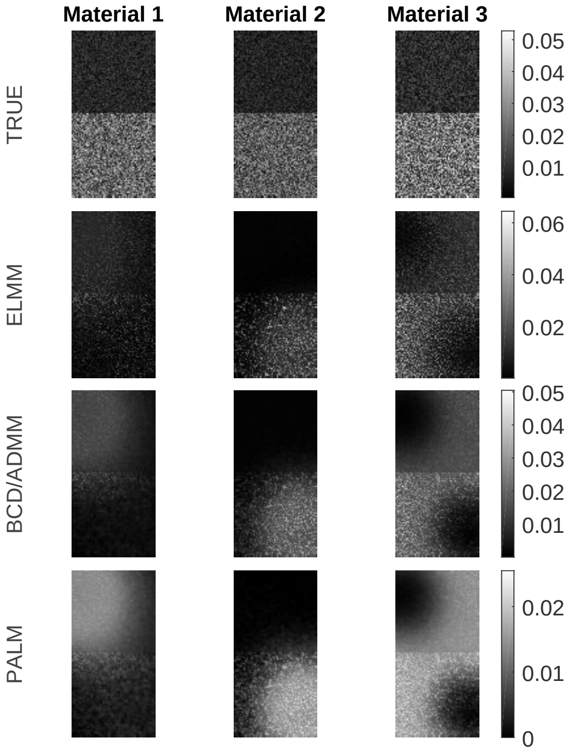


Figure 1.4.: Spatial distribution of the estimated variability with respect to each endmember, presented in terms of its energy ($\|\mathbf{dm}_{r,n}\|_2/\sqrt{L}$ for the rth endmember in the nth pixel).

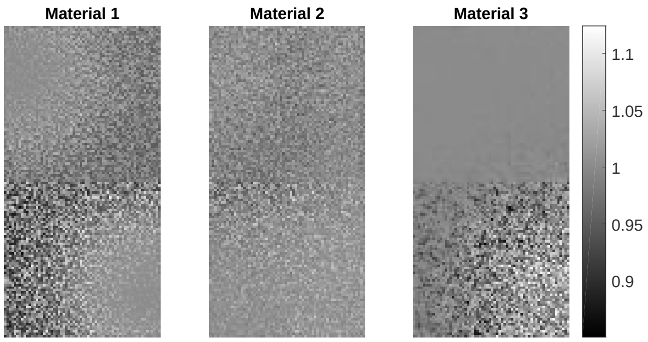
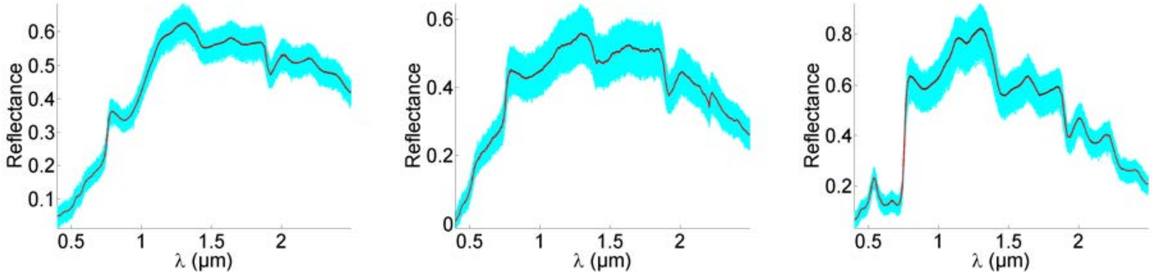


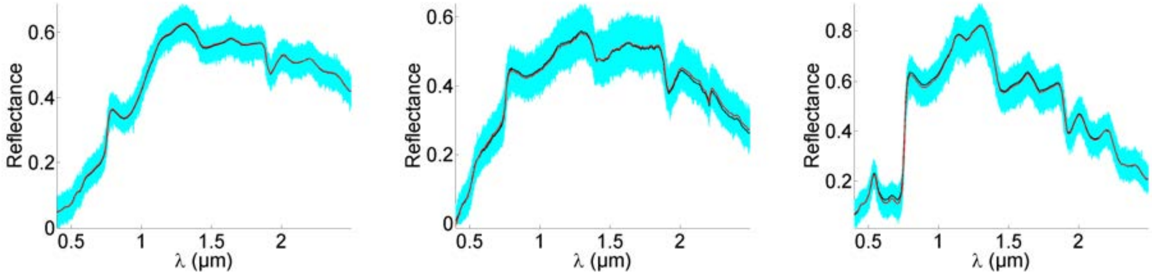
Figure 1.5.: Scaling coefficients recovered by the ELMM algorithm [Dru+16].



(a) Material 1 (ELMM)

(b) Material 2 (ELMM)

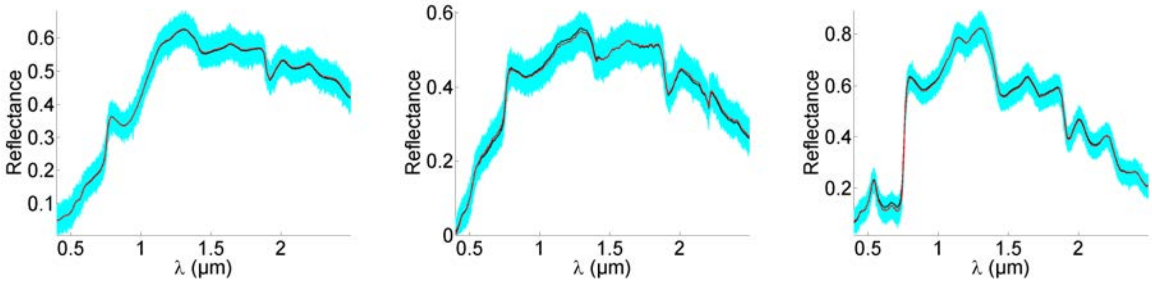
(c) Material 3 (ELMM)



(d) Material 1 (BCD/ADMM)

(e) Material 2 (BCD/ADMM)

(f) Material 3 (BCD/ADMM)



(g) Material 1 (PALM)

(h) Material 2 (PALM)

(i) Material 3 (PALM)

Figure 1.6.: Endmembers recovered from the synthetic dataset with $R = 3$ in absence of pure pixels. The estimated endmembers (red lines) are given with typical examples of the estimated variability (cyan lines). The VCA endmembers are given in black for comparison.

1.6. Experiments with real data

1.6.1. Description of the datasets

The proposed algorithm has been applied to the following real AVIRIS datasets[1].

- **Moffett scene:** acquired over Moffett Field (CA) in 1997, this 50 × 50 scene is partly composed of a lake and a coastal area. Unmixing procedures are conducted with 189 out of the the 224 available spectral bands due to the presence of water absorption bands. The scene has been unmixed with $R = 3$ endmembers based on prior studies conducted on this dataset [Dob+09; Hal+11; EDT11].
- **Cuprite scene:** the second scene is a 190 × 250 image extracted from the extensively studied Cuprite dataset, for which reference abundance maps are available from the literature. The unmixing procedure has been conducted with $R = 10$ endmembers based on the results obtained in prior studies [NB05b; MQ07]. Removing the water-absorption and low SNR bands leads to exploit 189 out of the 224 available spectral bands.

The parameters used for the proposed approach are identical to those used in the experiments with synthetic data (see Table 1.1). The values selected by a procedure similar to the one described in the previous section for α, β and γ for the BCD/ADMM algorithm (ν for the PALM algorithm) are given in Table 1.4.

1.6.2. Results

The unmixing performances reported in Table 1.4 lead to the following comments.

Moffett scene. The estimated abundance maps and endmembers are reported in Figs. 1.7 and 1.8, whereas the variability detected by the different algorithms is displayed in Figs. 1.9. The variability captured by the proposed methods 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 area [ADT11; FD15; Hal+11] are interpreted as variability in the proposed method, which tends to corroborate its consistency. For this specific scene, the ELMM does not lead to conclusive results in terms of variability energy, and in terms of the recovered scaling factors (see Fig. 1.10). In addition, the level of variability estimated by ELMM for the different endmembers suggests that some noise has been captured by the estimated perturbed endmembers (see e.g. Fig. 1.8(a)), which partly explains the lower RE and aSAM obtain with this approach. Note that the proposed method and ELMM do not require to consider a sophisticated non-linear model accounting for interactions between the different endmembers as in [ADT11; Hal+11; Dob+14b], which is an undeniable advantage. Conversely, all deviations from the LMM are contained in the variability components $\mathbf{dm}_{r,n}$ of the PLMM. We can also note that the variability peaks observed in Fig. 1.8 are a clear indication that several corrupted spectral bands (i.e., low SNR bands) have not been removed prior to the unmixing process.

---

[1] 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 1.4.: Simulation results on real data (RE $\times 10^{-4}$) $[(\alpha, \beta) = (2.5 \times 10^{-5}, 1.2 \times 10^{-6})$ for Moffett, $(\alpha, \beta) = (9 \times 10^{-2}, 5.2 \times 10^{-6})$ for Cuprite].

	Algorithm	RE	aSAM($\mathbf{Y}$) (°)	time (s)
Moffett ($R = 3$)	VCA/FCLS	2.50	7.31	**0.4**
	SISAL/FCLS	1.12	4.93	30
	FDNS	2.69	7.33	1
	AEB	6.25	6.93	10
	ELMM	**0.29**	**2.45**	64
	ssmdBCD/ADMM ($\gamma = 10^{-2}$)	0.45	2.65	102
	ssmdPALM ($\nu = 2 \times 10^{-1}$)	0.48	1.58	72
Cuprite ($R = 10$)	VCA/FCLS	3.69	1.39	**9.9**
	SISAL/FCLS	2.16	1.33	15
	FDNS	3.69	1.40	11
	AEB	0.40	0.90	615
	ELMM	0.15	0.49	2455
	ssmdBCD/ADMM ($\gamma = 10^{-1}$)	0.02	0.19	1749
	ssmdPALM ($\nu = 5 \times 10^{-1}$)	**1.53** $\times 10^{-2}$	0.03	2034

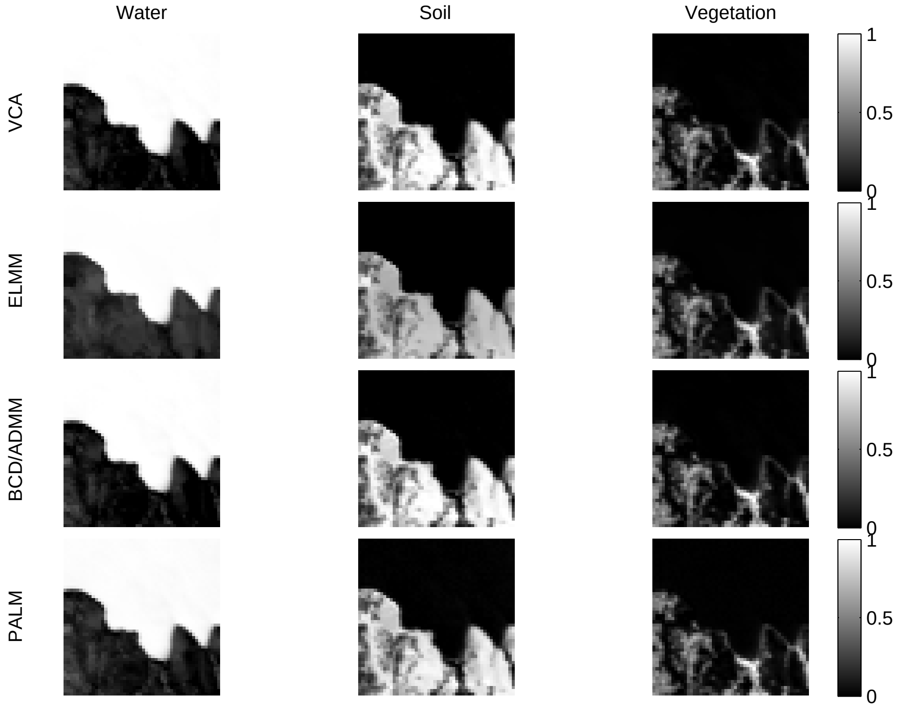
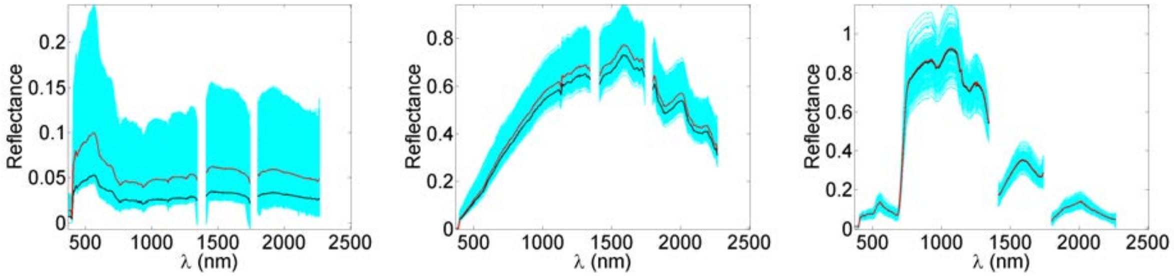
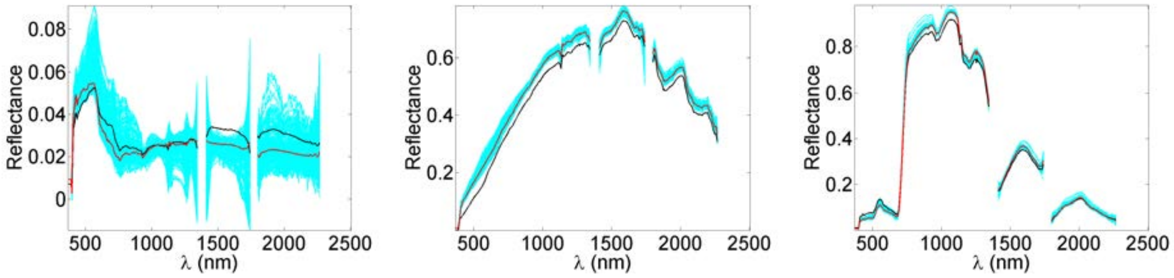


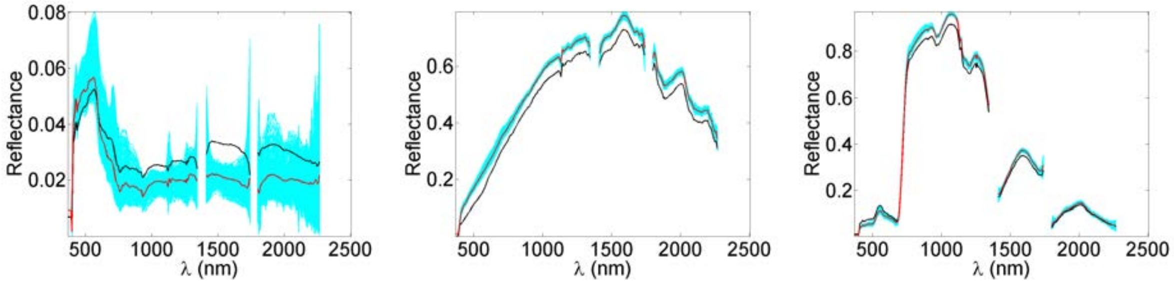
Figure 1.7.: Abundance maps estimated for the Moffett scene.



(a) Water (ELMM) (b) Vegetation (ELMM) (c) Soil (ELMM)



(d) Water (BCD/ADMM) (e) Vegetation (BCD/ADMM) (f) Soil (BCD/ADMM)



(g) Water (PALM) (h) Vegetation (PALM) (i) Soil (PALM)

Figure 1.8.: Endmembers estimated by ELMM, BCD/ADMM and PALM for the Moffett scene. Each endmember estimated by ELMM, BCD/ADMM and PALM (red lines) is plotted with the corresponding VCA endmember (in black lines), and typical examples of the estimated variability are given in cyan dotted lines.

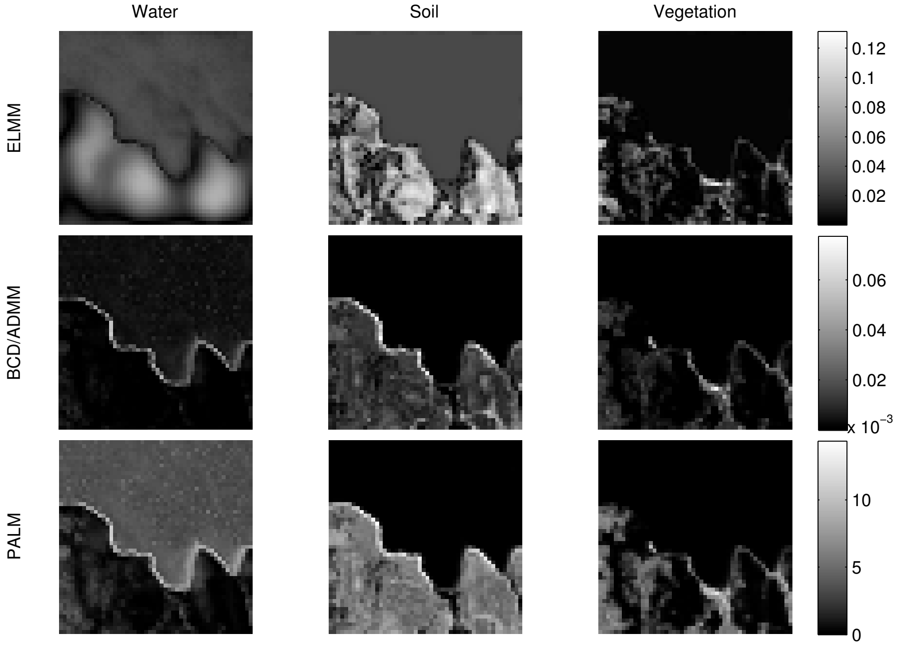


Figure 1.9.: Spatial distribution of the variability with respect to each endmember estimated for the Moffett dataset. The maps are presented in terms of the variability energy for visualization purpose ($\|\mathbf{dm}_{r,n}\|_2/\sqrt{L}$ for the rth endmember in the nth pixel). To allow a better comparison with the proposed methods, the variability captured by the ELMM is interpreted as the deviation of each corrupted spectral signature $\mathbf{m}_{r,n}$ (see (I.7)) from its (spatial) average $\bar{\mathbf{m}}_r$.

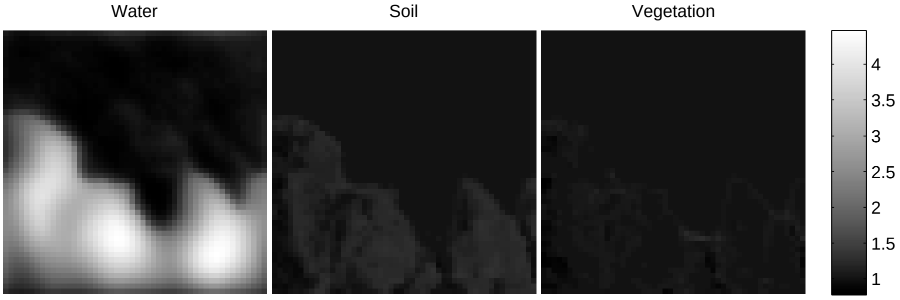


Figure 1.10.: Scaling coefficients recovered by the ELMM algorithm [Dru+16] from the Moffett dataset.

Cuprite scene. The results obtained for the Cuprite scene are reported in Figs. 1.11 to 1.16. Comparing our results with those of [NB05b], we visually found out that similar endmembers that were identified as different signatures by VCA for $R = 14$ [NB05b] are interpreted as multiple instances of single endmembers in our setting ($R = 10$). The identification is notably given in Fig. 1.11. Fig. 1.12 shows that the algorithms 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. The maps of the scaling factors recovered by ELMM, reported in Fig. 1.13, mainly exhibit variations in area where the other methods captured a significant variability contribution.

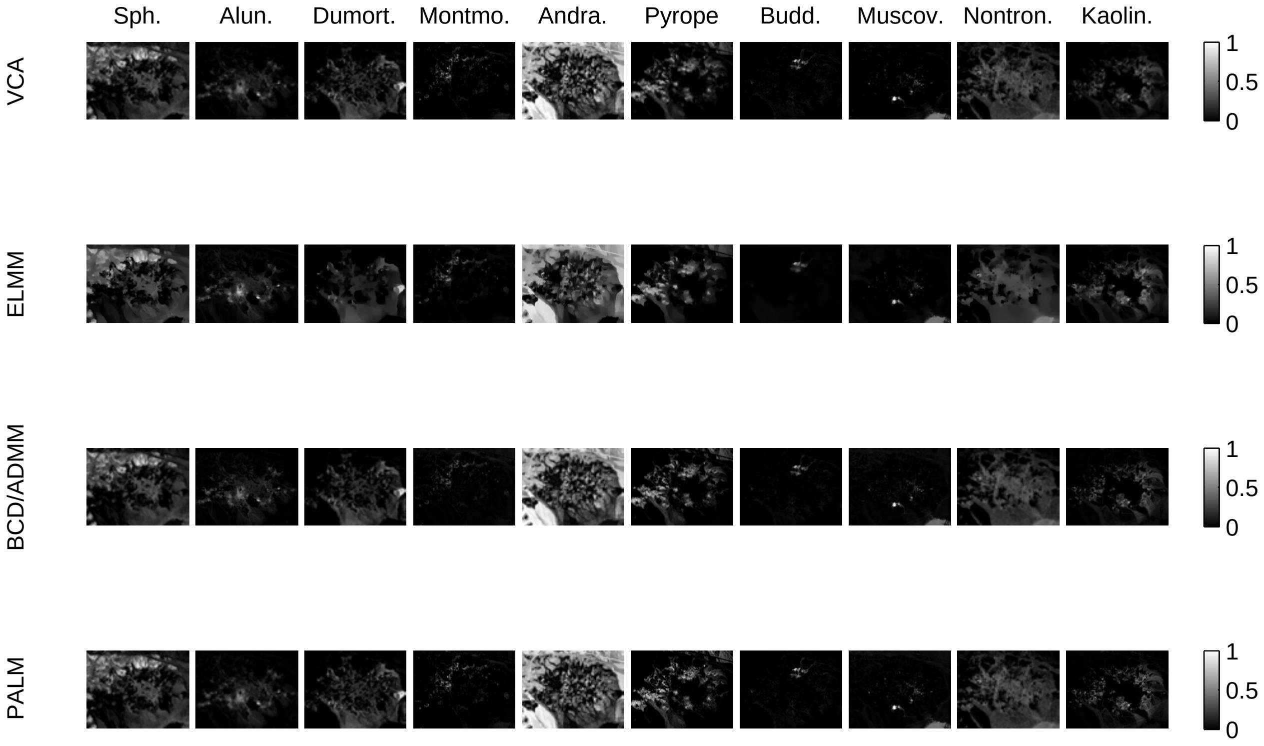


Figure 1.11.: Abundance maps estimated by ELMM, BCD/ADMM and PALM for the Cuprite scene. The given identification is based on a visual comparison with the results obtained in [NB05b] [Sphene, Alunite, Dumortierite, Montmorillonite, Andradite, Pyrope, Buddingtonite, Muscovite, Nontronite, Kaolinite].

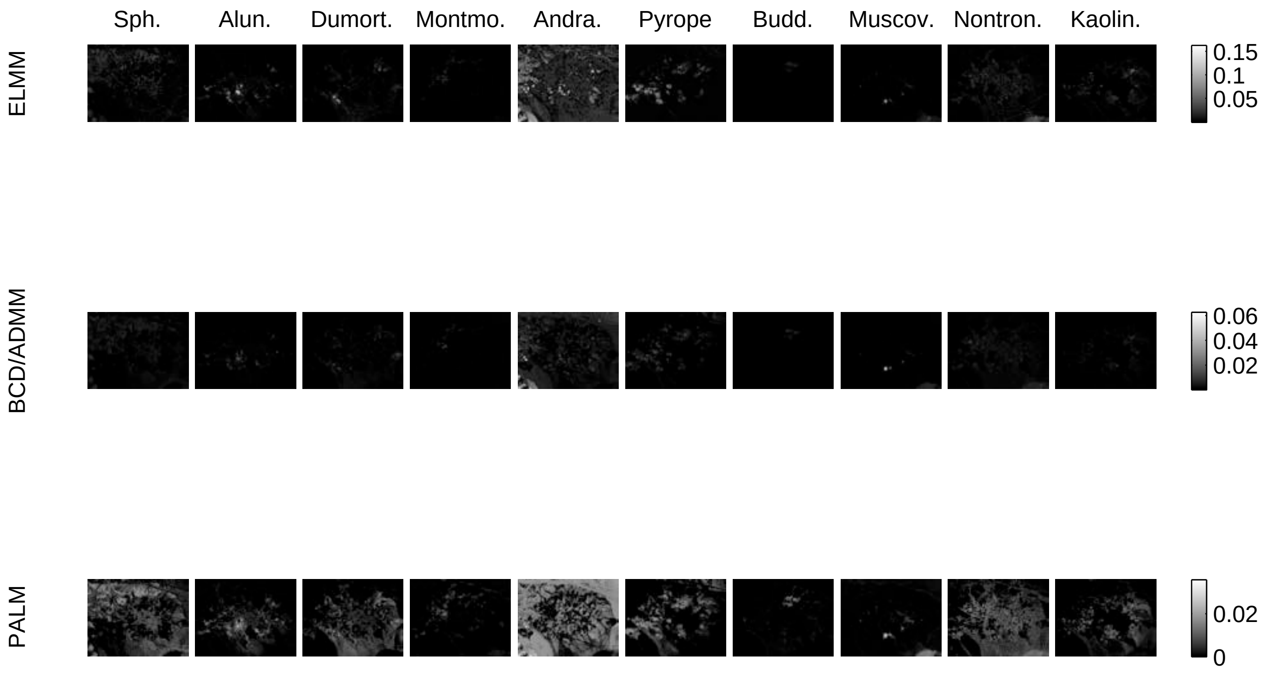


Figure 1.12.: Spatial distribution of the estimated variability for the Cuprite dataset, presented in terms of energy ($\|\mathbf{dm}_{r,n}\|_2/\sqrt{L}$ for the rth endmember in the nth pixel).

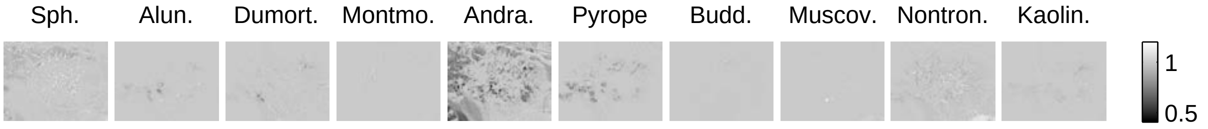
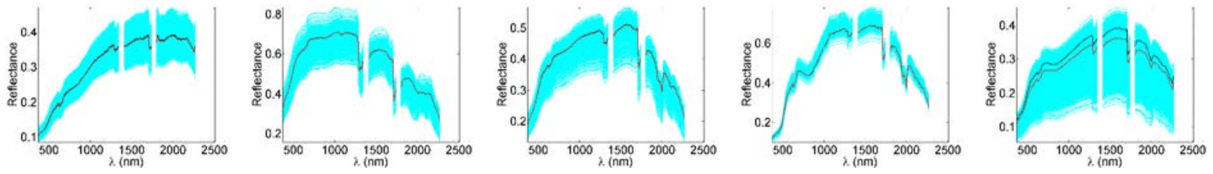
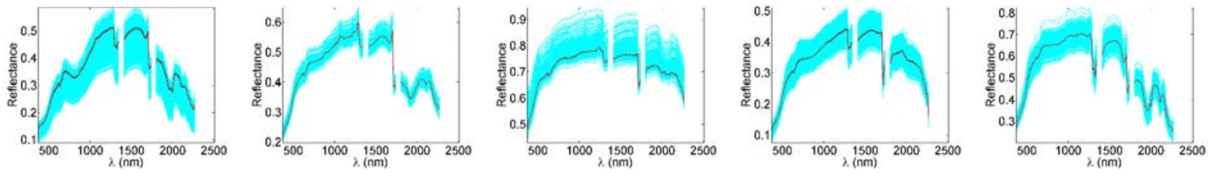


Figure 1.13.: Scaling coefficients recovered by the ELMM algorithm [Dru+16] from the Cuprite dataset.

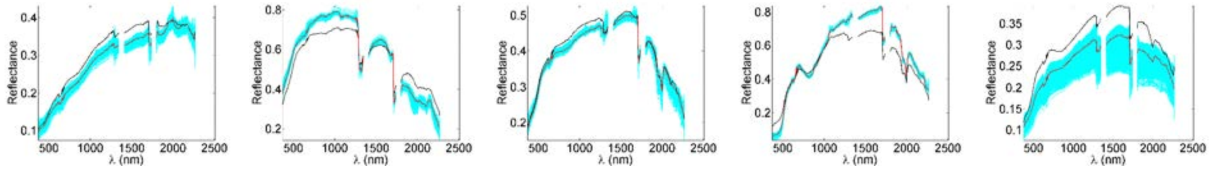


(a) Sphene (b) Alunite (c) Dumortierite (d) Montmorillonite (e) Andradite

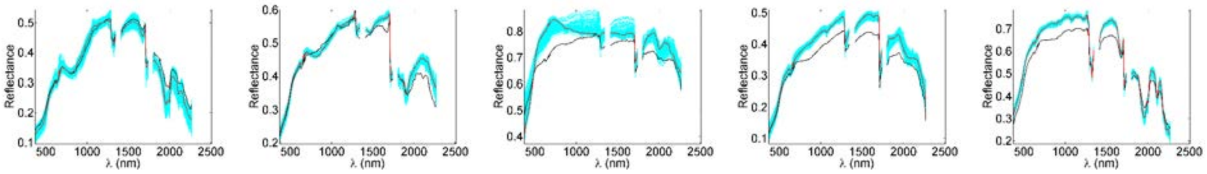


(f) Pyrope (g) Buddingtonite (h) Muscovite (i) Nontronite (j) Kaolinite

Figure 1.14.: Endmembers estimated by ELMM for the Cuprite scene. The given identification is based on a visual comparison with the results obtained in [NB05b].

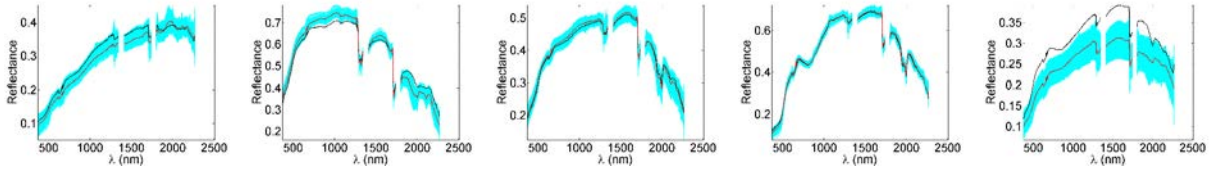


(a) Sphene (b) Alunite (c) Dumortierite (d) Montmorillonite (e) Andradite

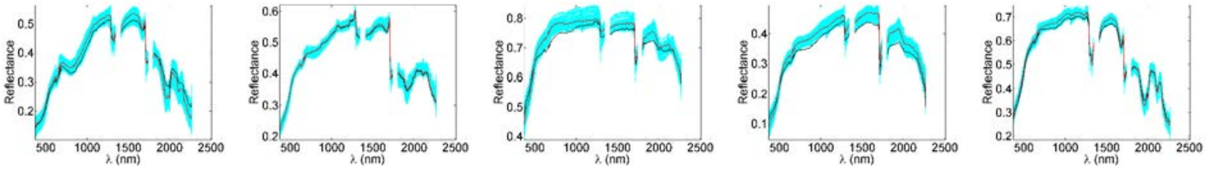


(f) Pyrope (g) Buddingtonite (h) Muscovite (i) Nontronite (j) Kaolinite

Figure 1.15.: Endmembers estimated by BCD/ADMM for the Cuprite scene. The given identification is based on a visual comparison with the results obtained in [NB05b].



(a) Sphene (b) Alunite (c) Dumortierite (d) Montmorillonite (e) Andradite



(f) Pyrope (g) Buddingtonite (h) Muscovite (i) Nontronite (j) Kaolinite

Figure 1.16.: Endmembers estimated by PALM for the Cuprite scene. The given identification is based on a visual comparison with the results obtained in [NB05b].

1.7. Conclusion and future work

This chapter introduced a new linear mixing model accounting for spatial endmember variability. The proposed model extends the classical LMM by including an additive spatially varying perturbation matrix that can capture endmember variabilities. The resulting unmixing problem was solved by alternating minimization of an appropriately regularized cost function, each minimization being performed by an appropriate ADMM algorithm. Another estimation strategy based on the PALM algorithm was also investigated. Simulations conducted on synthetic and real data enabled the interest of the proposed solution to be appreciated. Indeed, the recovered results compared favorably with state-of-the-art approaches while providing a relevant variability estimation.

The choice of the penalization parameters α, β and γ (or ν), performed by trying different values on a predefined grid of values, would deserve to be automated in the future. Finally, due to the significant number of unknown parameters, the proposed method is not intended to be applied to large images. This approach can be applied as a complementary tool to analyze 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 chapter is another interesting prospect.

With a first chapter focused on the representation of spatial endmember variability (i.e., within a single image), the next chapter addresses temporal endmember variability when unmixing a sequence of HS images – acquired over the same area at different time instants.

Main contributions. The contribution of this chapter lies in the introduction of an explicit parametric variability model. A block coordinate descent algorithm is first proposed to estimate the parameters of interest, which consists in solving strongly convex subproblems by ADMM. The proposed approach yields competitive reconstruction performance and consistent variability estimates on both synthetic and real data. The performance of the proximal algorithm [BST13; CPR16] is finally evaluated for variability estimation.

1.8. Conclusion (in French)

Un nouveau modèle de démélange a été introduit dans ce chapitre pour rendre compte de la variabilité spatiale des spectres purs. Ce modèle étend le modèle linéaire communément adopté, dans la mesure où il comporte un terme de perturbation additif spatialement variable pour représenter la variabilité spatiale des sources. Le problème de démélange obtenu a été résolu à l'aide de deux algorithmes de minimisation alternée, l'un basé sur l'algorithme ADMM, l'autre sur une descente proximale linéarisée. Les résultats des simulations conduites sur données synthétiques et réelles ont montré l'intérêt du modèle proposé pour rendre compte du phénomène de variabilité spatiale, notamment sur données réelles.

Le choix des paramètres de régularisation du modèle, effectué jusqu'ici en parcourant une grille de valeurs prédéfinies, nécessiterait le développement de procédures automatiques. Enfin, le nombre conséquent de degrés de liberté supplémentaires introduit par le modèle limite de fait son application à des images de taille modérée. Cette approche reste envisageable en tant qu'outil d'analyse complémentaire, pour une application à des images de taille limitée, *a priori* affectées par un niveau non-négligeable de variabilité. Enfin, réduire le coût calculatoire engendré par les approches considérées constitue une dernière perspective qu'il conviendrait d'approfondir.

Tandis que ce premier chapitre a permis de rendre compte de la variabilité spatiale des spectres purs au sein d'une unique image, le chapitre suivant propose de caractériser la variabilité temporelle des spectres purs lors du démélange d'une série multi-temporelle d'images hyperspectrales.

Contributions principales. Un modèle explicite de variabilité spatiale est proposé dans le cadre du démélange d'images hyperspectrales. Un permier algorithme d'optimisation, qui consiste à résoudre une séquence de sous-problèmes strictement convexes par la méthode des multiplicateurs de Lagrange (ADMM) dans un algorithme de descente par coordonnées, permet d'estimer les paramètres du modèle. Les performances de l'algorithme PALM [BST13; CPR16] sont également évaluées dans ce contexte. Ces deux algorithmes permettent d'obtenir des résultas de reconstruction compétitifs, tout en donnant une représentation satisfaisante de la variabilité.

CHAPTER 2

A Bayesian model accounting for endmember variability and abrupt spectral changes to unmix multitemporal hyperspectral images.

This chapter has been adapted from the submitted journal paper [TDT17a].

Contents

2.1. Introduction (in French)

En pratique, la forme et l'amplitude des signatures acquises sont particulièrement affectées par la variation des conditions d'acquisition, telles que l'illumination de la scène ou l'évolution naturelle du milieu observé. Ainsi, un spectre pur extrait pour un matériau donné peut présenter une variabilité significative d'une image à l'autre. À ce titre, le démélange simultané de plusieurs images hyperspectrales – acquises au-dessus d'une même scène à différents instants – présente un intérêt notable. En effet, l'utilisation simultanée de plusieurs images s'avère particulièrement pertinente en vue d'améliorer la qualité des résultats de démélange, notamment lorsque les matériaux observés sont correctement représentés non pas dans une seule, mais dans différentes images en fonction de leur date d'acquisition. Les résultats obtenus par de telles procédures doivent permettre d'améliorer les résultats par rapport à une analyse indépendante de chacune des images disponibles [Goe+13], tout en permettant de caractériser l'évolution temporelle des spectres purs.

Tandis que la variabilité spatiale des spectres purs a déjà fait l'objet d'un très grand nombre de publications (voir chapitre 1), un nombre comparativement plus faible de travaux est consacré à l'étude de la variabilité temporelle des sources dans une image hyperspectrale multi-dates. Dans le cadre du démélange hyperspectral, on peut noter que des séquences d'images hyperspectrales multi-temporelles ont été utilisées pour la détection de changements [EP15; Liu+16], ou dans la perspective d'améliorer les résultats de démélange en considérant des données multi-capteurs [YZP17]. Concernant la variabilité temporelle, les articles [Hal+15; HCJ16] comptent parmi les premières contributions liées plus particulièrement au problème de démélange. Bien que certaines approches de la littérature permettent d'exploiter une variation temporelle modérée de certains des paramètres du mélange [HCJ16; Hal+15], celles-ci ne rendent pas compte de variations spectrales abruptes, telles que générées par la présence de données aberrantes ou un intervalle de temps significatif entre deux acquisitions consécutives. De telles variations sont communément rencontrées lors de l'analyse de données réelles, en fonction de la date d'acquisition et de l'évolution naturelle de la scène (notamment en présence d'eau ou de végétation dans la scène). À moins d'être spécifiquement prises en compte, ces variations abruptes peuvent avoir un impact significatif sur le démélange d'images hyperspectrales.

Inspirés par les travaux [Hal+15; AMH15; CBR15] et dans le cadre d'une interprétation originale du problème de démélange, nous proposons de rendre simultanément compte de variations modérées des spectres purs – interprétées en termes de variabilité – et de variations spectrales abruptes – décrites comme des données aberrantes (résultant par exemple d'une variabilité importante des sources ou liées à la présence de non-linéarités). En particulier, la méthode proposée s'avère particulièrement adaptée à l'analyse de scènes pour lesquels les mêmes matériaux sont observés dans plusieurs images hyperspectrales acquises à des instants différents. Notre approche s'appuie sur des résultats de démélange obtenus pour une scène de référence issue de la séquence étudiée, utilisés ensuite comme point de départ pour démélanger l'ensemble des images de la séquence. Par ailleurs, les signatures identifiées dans chacune des images peuvent être considérées comme la réalisation de signatures de référence communes à l'ensemble des images, ce qui justifie l'utilisation d'une variante du modèle introduit au chapitre 1. À cet égard, les signatures associées à des matériaux qui n'apparaissent pas dans la scène de référence peuvent être interprêtées comme des données aberrantes vis-à-vis des signatures communes à l'ensemble de la séquence.

À la différence du chapitre précédent, nous introduisons un modèle bayésien visant à représenter simultanément la variabilité temporelle des sources et la présence éventuelle de données aberrantes lors du démélange d'images hyperspectrales multi-temporelles. La procédure de démélange est conduite à l'aide d'un algorithme de Monte-Carlo par chaîne de

Markov, qui permet de contruire des estimateurs des paramètres du modèle.

La suite de ce chapitre est organisée comme suit. Le modèle de mélange est introduit dans la section 2.3, tandis que le modèle bayésien associé est développé dans la section 2.4. Un échantillonneur de Gibbs hybride est étudié en section 2.5 afin d'estimer les paramètres du modèle considéré, qui comporte à la fois des variables discrètes (binaires) et continues. Les performances de l'approche proposée sont évaluées sur données synthétiques et réelles dans les sections 2.6 et 2.7, et sont comparées à celles obtenues par plusieurs algorithmes de la littérature (VCA/FCLS [NB05b; BF10], SISAL/FCLS [Bio09], RLMM [FD15] et le démélange en ligne du chapter 3). Les conclusions et perspectives de recherche sont enfin présentées dans la section 2.8.

2.2. Introduction

Varying acquisition conditions, such as illumination variations or the natural evolution of the scene, may significantly alter the shape and the amplitude of the acquired spectral signatures. Consequently, the endmembers extracted for a given material may vary significantly from an image to another. In this context, sequences of HS images acquired over the same area at different time instants can be of interest to exploit information redundancy between consecutive images (e.g., through features exhibiting moderate or smooth temporal variations [HMT15; Hal+16a]). Indeed, the use of several images is particularly appealing to improve the unmixing results when all the observed materials are not well represented in each image. The results obtained by such unmixing strategies are consequently expected to outperfom those relying on an individual analysis of each image [Goe+13].

Whereas spatial endmember variability has been extensively considered in the literature (see chapter 1), fewer works have considered the question of temporal variability in multitemporal HS (MTHS) images. MTHS images have been used to improve endmember unmixing results in [HCJ16; Hal+15; YZP17], or in the context of change detection in [EP15; Liu+16]. As for temporal variability, [Hal+15; HCJ16] are among the first to specifically address this issue in hyperspectral unmixing. Even though the approaches proposed in [HCJ16; Hal+15] allow smooth temporal variations of some of the mixture parameters to be exploited, they do not account for abrupt spectral changes, either due to outliers or to possibly significant time intervals between consecutive images. In practice, such situations can be reasonably expected, depending on the acquisition dates and possible climatic hazards, e.g., when vegetation or water is present in the observed scene. Unless specifically accounted for, this situation frequently observed in real datasets can have a significant impact on the recovered endmembers.

Inspired by [Hal+15; AMH15; CBR15] and based on an original interpretation of the unmixing problem under study, we propose to jointly account for smooth endmember variations – construed as temporal endmember variability – and abrupt changes interpreted as outliers (e.g., significant variability within a single image or presence of non-linearities). More precisely, the proposed analysis is focused on scenes in which mostly the same materials are expected to be observed from an image to another. Taken as a starting point, the results obtained from a reference scene can be reasonably extended to the whole MTHS image in this context. On the one hand, the endmembers identified in each single image can *in fine* be considered as time-varying instances of reference signatures shared by the different images, thus justifying the use of a modified version of the perturbed linear mixing model (PLMM) proposed in the previous chapter. This formulation notably allows smooth spectral variations occurring over time to be captured. On the other hand, the signatures corresponding to materials appearing in only a few images, inducing abrupt spectral changes, can be regarded as outliers with respect to the commonly shared endmembers.

This chapter studies a new Bayesian model allowing both the spectral variability and the presence of outliers to be considered in the unmixing of MTHS images. The resulting unmixing task is solved using a Markov chain Monte-Carlo (MCMC) method, which allows the posterior distribution of interest to be sampled and Bayesian estimators to be constructed.

The rest of this chapter is organized as follows. The mixing model considered is introduced in section 2.3, and the associated hierarchical Bayesian model is developed in section 2.4. Section 2.5 investigates an hybrid Gibbs sampler to solve the resulting problem, which involves both discrete and continuous variables. The performance of the proposed approach on synthetic and real data is assessed in sections 2.6 and 2.7. In particular, the results obtained with the proposed method are compared to those of the VCA/FCLS algorithm [NB05b; BF10], the SISAL/FCLS algorithm [Bio09], the algorithm associated with the robust LMM (RLMM) proposed in [FD15], the Dynamic spectral unmixing (DSU) introduced in [HCJ16] and the online unmixing algorithm introduced in the next chapter (see chapter 3). The comparison with this last method notably allows the robustness of the approach investigated in this chapter to be better appreciated. Conclusions and related research perspectives are finally given in section 2.8.

2.3. Problem statement

Throughout this chapter, we consider a sequence of HS images acquired at T different time instants over the same area. We further assume that the reference image is composed of R endmembers – where R is *a priori* known – which are likely to be shared with the $T-1$ remaining images. Since the signature of a given endmember can be reasonably expected to vary from an image to another, we propose to account for smooth endmember temporal variations via a modified version of the perturbed linear mixing model (PLMM) introduced in chapter 1. In practice, the PLMM model shows notable limitations when the vector $\mathbf{y}_{n,t}$ is affected by abrupt changes. Consequently, a new unmixing model is investigated to jointly account for endmember variability and abrupt changes possibly affecting MTHS images. To this end, an additional term $\mathbf{x}_{n,t}$ is considered to capture significant deviations from the PLMM, i.e., significant spatial variability or non-linearities within each image [FD15; AMH15]. The resulting observation model can thus be written

$$\mathbf{y}_{n,t} = \sum_{r=1}^{R} a_{r,n,t}\Big(\mathbf{m}_r + \mathbf{dm}_{r,t}\Big) + \mathbf{x}_{n,t} + \mathbf{b}_{n,t}, \ \forall n \in \{1,\dots,N\}, \ t \in \{1,\dots,T\} \tag{2.1}$$

where $\mathbf{y}_{n,t}$ denotes the nth image pixel within the image t, $\mathbf{m}_r$ is the rth endmember, $a_{r,n,t}$ is the proportion of the rth endmember in the nth pixel at time t, $\mathbf{dm}_{r,t}$ denotes the perturbation of the rth endmember at time t, and $\mathbf{x}_{n,t}$ denotes the contribution of outliers in the nth pixel at time t. Finally, $\mathbf{b}_{n,t}$ represents an additive noise resulting from the data acquisition and the modeling errors. In the following, the outliers $\mathbf{x}_{n,t}$ are assumed to be spatially sparse, which will allow their contribution to be discriminated from the noise. The proposed model (2.1) can be compactly written in matrix form as

$$\mathbf{Y}_t = (\mathbf{M} + \mathbf{dM}_t)\mathbf{A}_t + \mathbf{X}_t + \mathbf{B}_t \tag{2.2}$$

where $\mathbf{Y}_t = [\mathbf{y}_{1,t}, \ldots, \mathbf{y}_{N,t}]$ is an $L \times N$ matrix containing the pixels of the tth image, $\mathbf{M}$ denotes an $L \times R$ matrix containing the endmembers that are common to all the images of the sequence, $\mathbf{A}_t$ is an $R \times N$ matrix composed of the abundance vectors $\mathbf{a}_{n,t}$, $\mathbf{dM}_t$ is an $L \times R$ matrix whose columns contain the variability inherent to the tth image, $\mathbf{X}_t$ is an $L \times N$ matrix whose columns are the outliers present in the image t, and $\mathbf{B}_t$ is an $L \times N$ matrix accounting for the noise at time t. The constraints considered to reflect physical considerations are

$$\begin{aligned} &\mathbf{A}_t \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}_t^{\mathrm{T}} \mathbf{1}_R = \mathbf{1}_N \\ &\mathbf{M} \succeq \mathbf{0}_{L,R}, \quad \mathbf{M} + \mathbf{dM}_t \succeq \mathbf{0}_{L,R} \\ &\mathbf{X}_t \succeq \mathbf{0}_{L,N}, \quad \forall t \in \{1, \ldots, T\} \end{aligned} \tag{2.3}$$

where $\succeq$ denotes a term-wise inequality. More precisely, the abundance sum-to-one constraint will be relaxed in the following for pixels containing outliers, as detailed in section 2.4.2. Note that the outlier term $\mathbf{X}_t$ is intended to describe abrupt changes due to the apparition of new endmembers that were not present in the reference image. This justifies the corresponding non-negativity constraint, similar to the one imposed on the other endmembers. Given the model (2.2), the unmixing problem considered in this work consists in inferring the abundances $\mathbf{A}_t$, the endmembers $\mathbf{M}$, the variability $\mathbf{dM}_t$ and the outliers $\mathbf{X}_t$ from the observations $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_T]$. In the next section, this problem is tackled in a Bayesian framework to incorporate prior knowledge specific to each mixture parameter.

2.4. Bayesian model

2.4.1. Likelihood

Assuming the additive noise $\mathbf{b}_{n,t}$ is distributed according to a Gaussian distribution $\mathbf{b}_{n,t} \sim \mathcal{N}(\mathbf{0}_L, \sigma_t^2 \mathbf{I}_L)$, the observation model (2.2) leads to

$$\mathbf{y}_{n,t} \mid \mathbf{M}, \mathbf{dM}_t, \mathbf{A}_t, \mathbf{X}_t, \sigma_t^2 \sim \mathcal{N}\Big((\mathbf{M} + \mathbf{dM}_t)\mathbf{a}_{n,t} + \mathbf{x}_{n,t}, \sigma_t^2 \mathbf{I}_L\Big).$$

In addition, assuming prior independence between the images $\mathbf{Y}_t$, and between the pixels within each image, the likelihood function of the image sequence $\underline{\mathbf{Y}}$ is

$$p(\underline{\mathbf{Y}} \mid \mathbf{M}, \mathbf{d}\underline{\mathbf{M}}, \underline{\mathbf{A}}, \underline{\mathbf{X}}, \boldsymbol{\sigma}^2) \propto \prod_{t=1}^{T} (\sigma_t^2)^{-NL/2} \exp\Big(-\frac{1}{2\sigma_t^2} \|\mathbf{Y}_t - (\mathbf{M} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{X}_t\|_{\mathrm{F}}^2\Big) \tag{2.4}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm, and the underline notation stands for the overall set of the corresponding parameters. In the context under study, the prior independence assumption is justified by the image acquisition process, which has been conducted with possibly different sensors at different time instants. Appropriate prior distributions are associated to each parameter and hyperparameter in the following paragraphs, thereby defining the proposed hierarchical Bayesian model. To alleviate notations, dependencies with respect to constant parameters have been omitted.

2.4.2. Parameter priors

Abundances ($\underline{\mathbf{A}}$). To exploit possible information redundancy, we propose to promote smooth temporal variations of the abundances between successive time instants for pixels that are not classified as outliers. To this end, we introduce an abundance prior defined for $n \in \{1, \dots, N\}$ as

$$\mathbf{a}_{n,1} \mid \mathbf{x}_{n,t} = \mathbf{0}_L \sim \mathcal{U}_{\mathcal{S}_R} \tag{2.5}$$

$$\mathbf{a}_{n,t} \mid \mathbf{x}_{n,t} \neq \mathbf{0}_L \sim \mathcal{U}_{\widetilde{\mathcal{S}_R}},\ t \in \{1, \dots, T\} \tag{2.6}$$

$$p\big(\mathbf{a}_{n,t} \mid \mathbf{x}_{n,t} = \mathbf{0}_L, \underline{\mathbf{A}}_{\backslash\{\mathbf{a}_{n,t}\}}\big) \propto \exp\left\{-\frac{1}{2\varepsilon_n^2}\left([\mathscr{T}^1_{n,t} \neq \emptyset]\|a_{n,t} - a_{n,\tau^1_{n,t}}\|_2^2\right)\right\}\mathbb{1}_{\mathcal{S}_R}(\mathbf{a}_{n,t}),\ t \geq 2 \tag{2.7}$$

where $\mathcal{U}_{\mathcal{S}_R}$ denote the uniform distribution on the set $\mathcal{S}_R$, $\mathbb{1}_{\mathcal{S}_R}$ is the indicator function of the set $\mathcal{S}_R$, $[\mathscr{P}]$ denotes the Iverson bracket applied to the logical proposition $\mathscr{P}$, i.e.,

$$[\mathscr{P}] = \begin{cases} 1, & \text{if } \mathscr{P} \text{ is true;} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\mathcal{S}_R = \{\mathbf{x} \in \mathbb{R}^R \mid \forall i, x_i \geq 0 \text{ and } \mathbf{x}^\mathrm{T}\mathbf{1}_R = 1\}$$

$$\widetilde{\mathcal{S}_R} = \{\mathbf{x} \in \mathbb{R}^R \mid \forall i, x_i \geq 0 \text{ and } \mathbf{x}^\mathrm{T}\mathbf{1}_R \leq 1\}$$

$$\mathscr{T}^1_{n,t} = \{\tau < t \mid \mathbf{x}_{n,\tau} = 0\}, \quad \tau^1_{n,t} = \max_{\tau \in \mathscr{T}^1_{n,t}} \tau.$$

By convention, we define $\mathscr{T}^1_{n,1} = \emptyset$. To be more explicit about the prior defined in (2.5) to (2.7), consider an image at time t and a pixel n within this image which is not corrupted by outliers (i.e., $\mathbf{x}_{n,t} = \mathbf{0}_L$). For $t = 1$, a uniform distribution defined in the unit simplex is selected to reflect the absence of specific prior knowledge while accounting for the related constraints specified in (2.3). For $t > 1$, smooth variations of $\mathbf{a}_{n,t}$ are promoted *via* a one-dimensional Gaussian Markov field [Maz+15; HMT15], which penalizes the Euclidean distance between $\mathbf{a}_{n,t}$ and the abundance of the last outlier-free pixel, i.e., at time instant $\tau^1_{n,t}$. On the contrary, when outliers are present in the pixel (n,t) ($\mathbf{x}_{n,t} \neq \mathbf{0}_L$), the usual abundance sum-to-one constraint is relaxed ($\mathbf{a}_{n,t}^T\mathbf{1}_R \leq 1$) so that the prior allows cases in which the linear model does not exhaustively describe the data to be addressed. The joint abundance prior can be written

$$p(\underline{\mathbf{A}} \mid \underline{\mathbf{X}}) = \prod_{n=1}^{N}\Bigg[\prod_{\substack{j=1 \\ t_j : \mathbf{x}_{n,t_j} \neq \mathbf{0}_L}}^{J_n} p(\mathbf{a}_{n,t_j} \mid \mathbf{x}_{n,t_j} \neq \mathbf{0}_L)\Bigg]\Bigg[\prod_{\substack{i=1 \\ t_i : \mathbf{x}_{n,t_i} = \mathbf{0}_L}}^{I_n} p(\mathbf{a}_{n,t_i} \mid \mathbf{a}_{n,t_{i-1}}, \mathbf{x}_{n,t_i} = \mathbf{0}_L)\Bigg]$$

where $I_n = \sharp\{t : \mathbf{x}_{n,t} = \mathbf{0}_L\}$, $J_n = T - I_n$ and $\sharp$ denotes the cardinal operator.

Outliers and labels ($\underline{\mathbf{X}}$, $\mathbf{Z}$). Similarly to [AMH15], outliers are *a priori* assumed to be spatially sparse. This prior knowledge has been addressed in the literature *via* different approaches, either relying on the ℓ_1 penalty (such as the LASSO [Tib96]) or on mixtures of probability distributions involving a Dirac mass at zero and a continuous probability distribution [VS13] (such as Bernoulli-Laplace [DHT09] or Bernoulli-Gaussian distributions [KM82; Lav93; BC05; BDT11; CTC15]). In this chapter, we propose to assign the following spatial sparsity promoting prior to the outliers $\mathbf{x}_{n,t}$

$$p(\mathbf{x}_{n,t} \mid z_{n,t}, s_t^2) = (1 - z_{n,t})\delta(\mathbf{x}_{n,t}) + z_{n,t}\mathcal{N}_{\mathbb{R}_+^L}(\mathbf{0}_L, s_t^2) \tag{2.8}$$

where $\mathcal{N}_{\mathbb{R}_+^L}$ denotes a Gaussian distribution truncated to the set $\mathbb{R}_+^L$, and the latent variables $z_{n,t}$ represent the outliers' support ($z_{n,t} = 1$ if an outlier is present in the corresponding pixel, 0 otherwise).

The prior (2.8) allows outliers to be *a priori* described by a truncated Gaussian distribution when $z_{n,t} = 1$, since the outliers $\mathbf{x}_{n,t}$ are assumed to mainly result from the apparition of new endmembers (i.e., that were not present in the reference image). With this context in mind, we further propose to promote spatial correlations between the outliers' support, since new materials are likely to appear in several contiguous pixels. The binary label maps $\mathbf{z}_t \in \mathbb{R}^N$ ($t = 1, \dots, T$) are consequently modeled as Ising-Markov random fields [EDT11; Alt+14; HMT15], for which the Hammersley-Clifford theorem yields

$$\mathbb{P}[\mathbf{z}_t = \mathbf{z}_t^0 \mid \beta_t] = \frac{1}{C(\beta_t)} \exp\left[\beta_t \sum_{n=1}^{N} \sum_{k \in \mathcal{V}(n)} \delta(z_{n,t}^0 - z_{k,t}^0)\right]$$

where $C(\beta_t)$ is the partition function [Gio10] and $\mathcal{V}(n)$ denotes the 4-neighbourhood of the pixel n. Under appropriate prior independence assumptions with respect to the labels $\mathbf{z}_t$ and the outliers $\mathbf{x}_{n,t}$, we obtain

$$\begin{aligned} p(\underline{\mathbf{X}} \mid \mathbf{Z}, \mathbf{s}^2) &= \prod_{n,t} p(\mathbf{x}_{n,t} \mid z_{n,t}, s_t^2) \\ p(\mathbf{Z} \mid \boldsymbol{\beta}) &= \prod_{t=1}^{T} p(\mathbf{z}_t \mid \beta_t) \end{aligned}$$

with $\mathbf{Z} \in \mathbb{R}^{N \times T}$ and $\boldsymbol{\beta} \in \mathbb{R}^T$. Note that the prior (2.8) leads to the following result, which will be useful to sample the label maps in section 2.5.1

$$\begin{aligned} p(\mathbf{x}_{n,t} \mid \mathbf{z}_{\backslash n,t}, s_t^2, \beta_t) &= \sum_{i=0}^{1} p(\mathbf{x}_{n,t} \mid z_{n,t} = i, s_t^2)\mathbb{P}[z_{n,t} = i \mid \mathbf{z}_{\backslash n,t}, \beta_t] \\ &= (1 - \omega_{n,t})\delta(\mathbf{x}_{n,t}) + \omega_{n,t}\mathcal{N}_{\mathbb{R}_+^L}(\mathbf{0}_L, s_t^2\mathbf{I}_L) \end{aligned} \tag{2.9}$$

where $\mathbf{z}_{\backslash n,t}$ denotes the label map $\mathbf{z}_t$ whose nth entry has been removed, and

$$\omega_{n,t} = \mathbb{P}[z_{n,t} = 1 \mid \mathbf{z}_{\backslash n,t}, \beta_t] = \frac{1}{C} \exp\Big[\beta_t \sum_{k \in \mathcal{V}(n)} \delta(1 - z_{k,t})\Big] \tag{2.10}$$

$$C = \exp\Big[\beta_t \sum_{k \in \mathcal{V}(n)} \delta(1 - z_{k,t})\Big] + \exp\Big[\beta_t \sum_{k \in \mathcal{V}(n)} \delta(0 - z_{k,t})\Big]. \tag{2.11}$$

Remark 2.1. An asymmetric prior could have been used to represent the fact that outliers are *a priori* less likely to be observed, e.g., by considering an Ising Markov random field with a nonzero external magnetic field. To this end, the prior can be defined as

$$\mathbb{P}[\mathbf{z}_t = \mathbf{z}_t^0 \mid \beta_t, \xi_t] = \frac{1}{C(\beta_t)} \exp\Big[\beta_t \sum_{n=1}^{N} \sum_{k \in \mathcal{V}(n)} \delta(z_{n,t}^0 - z_{k,t}^0) + \xi_t z_{n,t}^0\Big]$$

$$\mathbb{P}[z_{n,t} = 1 \mid \mathbf{z}_{\backslash n,t}, \beta_t, \xi_t] = \frac{1}{C} \exp\Big[\log \xi_t + \beta_t \sum_{k \in \mathcal{V}(n)} \delta(1 - z_{k,t})\Big]$$

$$C = \exp\Big[\log \xi_t + \beta_t \sum_{k \in \mathcal{V}(n)} \delta(1 - z_{k,t})\Big] + \exp\Big[\beta_t \sum_{k \in \mathcal{V}(n)} \delta(0 - z_{k,t})\Big]$$

where $\xi_t \in \mathbb{R}_+^*$ for $t \in \{1, \dots, T\}$. The parameter ξ_t, set *a priori* or estimated from the data, directly controls the probability of observing outliers in the absence of any spatial regularization term. However, the partition function $C(\beta_t)$ does not admit an analytical expression in this setting, hence the need to resort to a different approach to estimate the granularity parameter β_t (see for instance [Per+13]).

Endmembers (M). The endmembers can be *a priori* assumed to live in a subspace of dimension $K = R - 1$ [Dob+09], which can be determined in practice by a principal component analysis (PCA) or a robust PCA (rPCA) [Can+09]. The PCA applied to the original data $\underline{\mathbf{Y}}$ leads to a decomposition which can be expressed as [Dob+09]

$$\mathbf{m}_r = \mathbf{U}\mathbf{e}_r + \check{\mathbf{y}}, \quad \check{\mathbf{y}} = (\mathbf{I}_L - \mathbf{U}\mathbf{U}^{\mathrm{T}})\bar{\mathbf{y}}, \quad \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_K \tag{2.12}$$

where $\mathbf{U}$ denotes a basis of the subspace of dimension K, and $\bar{\mathbf{y}}$ denotes the average spectral signature obtained from $\underline{\mathbf{Y}}$. The projected endmembers $\mathbf{e}_r$ are then assigned the following truncated Gaussian prior, which ensures the non-negativity of the endmembers $\mathbf{m}_r$

$$\mathbf{e}_r \sim \mathcal{N}_{\mathcal{E}_r}(\mathbf{0}_K, \xi \mathbf{I}_K), \text{ for } r \in \{1, \dots, R\} \tag{2.13}$$

with

$$\mathcal{E}_r = [e_{1,r}^-, e_{1,r}^+] \times \ldots \times [e_{K,r}^-, e_{K,r}^+]$$

$$e_{k,r}^- = \max_{\ell \in \mathcal{U}_k^+} \Big(-\frac{\check{y}_\ell + \sum_{j \neq k} u_{\ell,j} e_{j,r}}{u_{\ell,k}} \Big)$$

$$e_{k,r}^+ = \min_{\ell \in \mathcal{U}_k^-} \Big(-\frac{\check{y}_\ell + \sum_{j \neq k} u_{\ell,j} e_{j,r}}{u_{\ell,k}} \Big)$$

$$\mathcal{U}_k^- = \{r \in \{1,\dots,R\} : u_{k,r} < 0\}, \quad \mathcal{U}_k^+ = \{r \in \{1,\dots,R\} : u_{k,r} > 0\}.$$

Endmember variability ($\mathbf{d\underline{M}}$). To promote smooth temporal variations of the endmember variability while accounting for the term-wise non-negativity of the observed endmembers (i.e., $\mathbf{m}_r + \mathbf{dm}_{r,t} \geq \mathbf{0}_{L,R}$), we propose to assign the following prior to the variability vectors $\mathbf{dm}_{r,t}$

$$\mathbf{dm}_{r,1} \mid \mathbf{m}_r \sim \mathcal{N}_{\mathcal{I}_r}(\mathbf{0}_L, \nu \mathbf{I}_L), \text{ for } r = 1,\dots,R \tag{2.14}$$

$$dm_{\ell,r,t} \mid m_{\ell,r}, dm_{\ell,r,(t-1)}, \psi_{\ell,r}^2 \sim \mathcal{N}_{\mathcal{I}_{\ell,r}}\left(dm_{\ell,r,(t-1)}, \psi_{\ell,r}^2\right), \text{ for } \begin{array}{l} r = 1,\dots,R \\ \ell = 1,\dots,L \\ t = 1,\dots,T \end{array} \tag{2.15}$$

and $\mathcal{I}_{\ell,r} = [-m_{\ell,r}, +\infty)$. Assuming prior independence between the variability vectors finally leads to the following joint prior

$$p(\mathbf{d\underline{M}} \mid \mathbf{M}, \boldsymbol{\Psi}^2) = \prod_{r=1}^{R} \left[p(\mathbf{dm}_{r,1} \mid \mathbf{m}_r) \prod_{t=2}^{T} p(\mathbf{dm}_{r,t} \mid \mathbf{m}_r, \mathbf{dm}_{r,(t-1)}, \boldsymbol{\psi}_r^2) \right]$$

where $\boldsymbol{\Psi}^2 \in \mathbb{R}^{L \times R}$ contains the variability variances $\psi_{\ell,r}^2$.

Noise variances ($\boldsymbol{\sigma}^2$). A non-informative inverse-gamma conjugate prior is selected for the noise variance

$$\sigma_t^2 \sim \mathcal{IG}(a_\sigma, b_\sigma), \ \forall t \in \{1,\dots,T\} \tag{2.16}$$

with $a_\sigma = b_\sigma = 10^{-3}$ in order to ensure a weakly informative prior. Assuming the noise variances are *a priori* independent for different time instants finally leads to

$$p(\boldsymbol{\sigma}^2) = \prod_t p(\sigma_t^2)$$

where $\boldsymbol{\sigma}^2 \in \mathbb{R}^R$ contains the noise variances σ_t^2.

2.4.3. Hyperparameters

In order to complete the description of the proposed hierarchical Bayesian model, we consider the following generic priors for the different hyperparameters.

(i) Non-informative conjugate inverse-gamma priors have been adopted for the variability variances $\boldsymbol{\Psi}^2 \in \mathbb{R}^{L \times R}$ and the outlier variances $\mathbf{s}^2 \in \mathbb{R}^R$, i.e., for $(\ell, r, t) \in \{1,\dots,L\} \times \{1,\dots,R\} \times \{1,\dots,T\}$

$$\psi_{\ell,r}^2 \sim \mathcal{IG}(a_\Psi, b_\Psi), \quad s_t^2 \sim \mathcal{IG}(a_\mathbf{s}, b_\mathbf{s}) \tag{2.17}$$

where $\mathcal{IG}(a_\Psi, b_\Psi)$ denotes the inverse gamma distribution and $a_\Psi = b_\Psi = a_\mathbf{s} = b_\mathbf{s} = 10^{-3}$ (in order to obtain flat distributions). Appropriate independence assumptions lead to

$$p(\boldsymbol{\Psi}^2) = \prod_{\ell,r} p(\psi^2_{\ell,r}), \quad p(\mathbf{s}^2) = \prod_t p(s^2_t)$$

where $\mathbf{s}^2 \in \mathbb{R}^T$ is composed of the outlier variances s^2_t.

(ii) Previous studies have shown that the granularity parameter of a Potts-Markov random field (*a fortiori* of an Ising-Markov random field) can be reasonably constrained to the interval $[0,2]$ [Gio11]. Consequently, a uniform prior on the interval $[0,2]$ has been adopted

$$\beta_t \sim \mathcal{U}_{[0,2]}, \ \forall t \in \{1,\dots,T\}. \tag{2.18}$$

Assuming the granularity parameters are *a priori* independent for different time instants finally yields

$$p(\boldsymbol{\beta}) = \prod_t p(\beta_t)$$

where $\boldsymbol{\beta} \in \mathbb{R}^T$ is composed of the granularity parameters β_t.

The resulting hierarchical Bayesian model is summarized in the directed acyclic graph of Fig. 2.1, where fixed parameters appear in boxes.

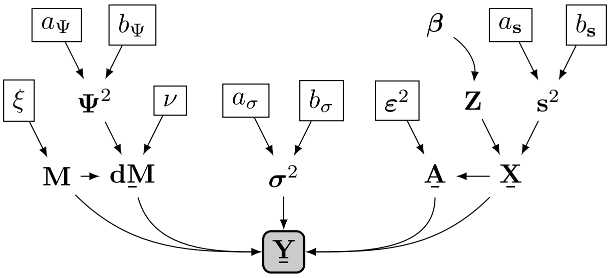


Figure 2.1.: Directed acyclic graph associated with the proposed Bayesian model (fixed parameters appear in boxes).

2.4.4. Joint posterior distribution

According to Bayes' theorem, the joint posterior distribution of the parameters of interest Θ is given by

$$\begin{aligned} p(\Theta \mid \underline{\mathbf{Y}}) \propto p(\underline{\mathbf{Y}} \mid \Theta) p(\underline{\mathbf{A}} \mid \underline{\mathbf{X}}) p(\underline{\mathbf{X}} \mid \mathbf{Z}, \mathbf{s}^2) p(\mathbf{s}^2) p(\mathbf{Z} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta}) \\ \times p(\mathbf{d\underline{M}} \mid \mathbf{M}, \boldsymbol{\Psi}^2) p(\mathbf{M}) p(\boldsymbol{\Psi}^2) p(\boldsymbol{\sigma}^2). \end{aligned} \tag{2.19}$$

where $\Theta = \{\Theta_p, \Theta_h\}$ and

$$\Theta_p = \{\mathbf{M}, \mathbf{d\underline{M}}, \underline{\mathbf{A}}, \underline{\mathbf{X}}, \boldsymbol{\sigma}^2, \mathbf{Z}\}, \ \Theta_h = \{\boldsymbol{\Psi}^2, \mathbf{s}^2, \boldsymbol{\beta}\}$$

denote the parameters and hyperparameters to be inferred. The complexity of the posterior density (2.19) does not allow maximum *a posteriori* (MAP) or minimum mean square (MMSE) estimators to be easily derived. For instance, the optimization problem associated with the determination of the MAP estimator of Θ is difficult to be solved, since the negative log-posterior is non-convex and parameterized by continuous and discrete variables. An MCMC method is consequently adopted to sample the posterior (2.19), thus allowing appropriate estimators of the parameters of interest to be computed from the generated samples.

2.5. Hybrid Gibbs sampler

This section studies an hybrid Gibbs sampler, which is guaranteed to produce samples asymptotically distributed according to the target distribution (2.19). This sampler – described in Algo. 4 – consists in generating samples distributed according to the conditional distribution of each parameter of interest. Section 2.5.2 introduces the proposed sampling method, whereas the conditional distributions of all the parameters of interest are detailed in the following paragraphs.

Algorithm 4: Proposed hybrid Gibbs sampler.

```
Input: $N_{\text{bi}}$, $N_{\text{MC}}$, $\Theta^{(0)}$, $\xi$, $a_\Psi$, $b_\Psi$, $a_{\mathbf{s}}$, $b_{\mathbf{s}}$, $a_\sigma$, $b_\sigma$, $\nu$, $\varepsilon^2$.
for $q = 1$ to $N_{MC}$ do
    for $(n,t) = (1,1)$ to $(N,T)$ do
        Draw $\mathbf{a}_{n,t}^{(q)} \sim p\Big(\mathbf{a}_{n,t} \mid \mathbf{y}_{n,t}, \Theta_{\backslash\{\mathbf{a}_{n,t}\}}\Big)$ ;
    for $r = 1$ to $R$ do
        Draw $\mathbf{e}_r^{(q)} \sim p\Big(\mathbf{e}_r \mid \underline{\mathbf{Y}}, \Theta_{\backslash\{\mathbf{e}_r\}}\Big)$ ;
    for $t = 1$ to $T$ do
        Draw $\mathbf{dM}_t^{(q)} \sim p\Big(\mathbf{dM}_t \mid \mathbf{Y}_t, \Theta_{\backslash\{\mathbf{dM}_t\}}\Big)$ ;
    for $(n,t) = (1,1)$ to $(N,T)$ do
        Draw $z_{n,t}^{(q)} \sim \mathbb{P}\Big[z_{n,t} \mid \mathbf{y}_{n,t}, \Theta_{\backslash\{z_{n,t}\}}\Big]$ ;
        Draw $\mathbf{x}_{n,t}^{(q)} \sim p\Big(\mathbf{x}_{n,t} \mid \Theta_{\backslash\{\mathbf{x}_{n,t}\}}\Big)$ ;
    for $t = 1$ to $T$ do
        Draw $s_t^{2(q)} \sim p\Big(s_t^2 \mid \Theta_{\backslash\{s_t^2\}}\Big)$ ;
    for $t = 1$ to $T$ do
        Draw $\sigma_t^{2(q)} \sim p\Big(\sigma_t^2 \mid \Theta_{\backslash\{\sigma_t^2\}}\Big)$ ;
    for $(\ell,r) = (1,1)$ to $(L,R)$ do
        Draw $\psi_{\ell,r}^{2(q)} \sim p\Big(\psi_{\ell,r}^2 \mid \Theta_{\backslash\{\psi_{\ell,r}^2\}}\Big)$ ;
    Sample the granularity parameters $\beta^{(q)}$, see §2.5.1 ;
Result: $\Big\{\Theta^{(q)}\Big\}_{q=1}^{N_{\text{MC}}}$.
```

2.5.1. Parameter sampling

Sampling the abundances ($\underline{\mathbf{A}}$). The likelihood function (2.4) combined with the prior given in section 2.4.2 leads to the following conditional distribution

$$\mathbf{a}_{n,t} \mid \mathbf{y}_{n,t}, z_{n,t} = 0, \Theta_{\backslash\{\mathbf{a}_{n,t}\}} \sim \mathcal{N}_{\mathcal{S}_R}(\boldsymbol{\mu}_{n,t}^{(\mathbf{A})}, \boldsymbol{\Lambda}_{n,t}) \tag{2.20}$$

$$\mathbf{a}_{n,t} \mid \mathbf{y}_{n,t}, z_{n,t} = 1, \Theta_{\backslash\{\mathbf{a}_{n,t}\}} \sim \mathcal{N}_{\widetilde{\mathcal{S}_R}}(\boldsymbol{\mu}^{(\mathbf{A})}_{n,t}, \boldsymbol{\Lambda}_{n,t}) \tag{2.21}$$

$$\boldsymbol{\Lambda}^{-1}_{n,t} = \frac{1}{\sigma_t^2}\mathbf{M}_t^{\mathrm{T}}\mathbf{M}_t + \frac{1}{\varepsilon_n^2}\big([\mathscr{T}^1_{n,t} \neq \emptyset] + [\mathscr{T}^2_{n,t} \neq \emptyset]\big)\mathbf{I}_R, \quad \mathbf{M}_t \triangleq \mathbf{M} + \mathbf{dM}_t \tag{2.22}$$

$$\boldsymbol{\mu}^{(\mathbf{A})}_{n,t} = \boldsymbol{\Lambda}_{n,t}\left[\frac{1}{\sigma_t^2}\mathbf{M}_t^{\mathrm{T}}(\mathbf{y}_{n,t} - \mathbf{x}_{n,t}) + \frac{1}{\varepsilon_n^2}\big([\mathscr{T}^1_{n,t} \neq \emptyset]\mathbf{a}_{n,\tau^1_{n,t}} + [\mathscr{T}^2_{n,t} \neq \emptyset]\mathbf{a}_{n,\tau^2_{n,t}}\big)\right] \tag{2.23}$$

where $\mathcal{N}_{\mathcal{S}_R}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ denotes a Gaussian distribution truncated to the set $\mathcal{S}_R$ and

$$\mathscr{T}^2_{n,t} = \{\tau > t \mid z_{n,\tau} = 0\}, \quad \tau^2_{n,t} = \min_{\tau \in \mathscr{T}^2_{n,t}} \tau \tag{2.24}$$

with the convention $\mathscr{T}^2_{n,T} = \emptyset$. Samples distributed according to the above truncated multivariate Gaussian distributions can be generated by the Gibbs sampler described in [DT07, section IV.B.] [AMD14], by an Hamiltonian Monte-Carlo procedure [ADT14; PP14] or by the general method recently proposed in [Bot16]. In this work, the Gibbs sampler [DT07, section IV.B.] has been adopted to sample the parameters of interest. Note that the abundance vectors $\mathbf{a}_{n,t}$ can be sampled in parallel to accelerate the algorithm.

Sampling the projected endmembers (E). Combining (2.4) and the prior given in section 2.4.2, the conditional distributions of the projected endmembers $\mathbf{e}_r$, for $r \in \{1, \dots, R\}$, can be shown to be

$$\mathbf{e}_r \mid \underline{\mathbf{Y}}, \Theta_{\backslash\{\mathbf{e}_r\}} \sim \mathcal{N}_{\mathscr{E}_r}(\boldsymbol{\mu}^{(\mathbf{E})}_r, \boldsymbol{\Lambda}_r) \tag{2.25}$$

with, for $k \in \{1, \dots, K\}$

$$\mathscr{E}_r = [c_{1,r}, d_{1,r}] \times \ldots \times [c_{K,r}, d_{K,r}] \tag{2.26}$$

$$c_{k,r} = \max_{\ell \in \mathcal{U}_k^+} \left(-\frac{\check{y}_\ell + \sum_{j \neq k} u_{\ell,j} e_{j,r} + b_{\ell,r}}{u_{\ell,k}}\right) \tag{2.27}$$

$$d_{k,r} = \min_{\ell \in \mathcal{U}_k^-} \left(-\frac{\check{y}_\ell + \sum_{j \neq k} u_{\ell,j} e_{j,r} + b_{\ell,r}}{u_{\ell,k}}\right) \tag{2.28}$$

$$b_{\ell,r} = \min\Big\{0, \min_t (dm_{\ell,r,t})\Big\}, \quad \boldsymbol{\Lambda}_r^{-1} = \left[\frac{1}{\xi} + \sum_{n,t} \frac{a^2_{r,n,t}}{\sigma_t^2}\right]\mathbf{I}_{R-1} \tag{2.29}$$

$$\boldsymbol{\mu}^{(\mathbf{E})}_r = \boldsymbol{\Lambda}_r \mathbf{U}^{\mathrm{T}}\left[\sum_{t,n} \frac{1}{\sigma_t^2}\Big(\mathbf{y}_{n,t} - \mathbf{x}_{n,t} - \mathbf{dM}_t\mathbf{a}_{n,t} - \check{\mathbf{y}} a_{r,n,t} - \sum_{j \neq r} a_{j,n,t}\mathbf{m}_j\Big) a_{r,n,t}\right]. \tag{2.30}$$

Samples distributed according to the above truncated Gaussian distributions can be efficiently generated with the algorithm described in [Cho11].

Sampling the variability (d$\underline{\mathbf{M}}$). Similarly, the likelihood function (2.4) and the prior given in section 2.4.2 lead to

$$dm_{\ell,r,t} \sim \mathcal{N}_{[-m_{\ell,r},+\infty)}(\mu^{(\mathbf{dM})}_{\ell,r,t}, \eta^2_{\ell,r,t}) \tag{2.31}$$

with

$$\frac{1}{\eta^2_{\ell,r,t}} = \frac{1}{\sigma_t^2}\sum_n a^2_{r,n,t} + \frac{1}{\nu}[t=1] + \frac{1}{\psi^2_{\ell,r}}\Big(1+\big[1<t<T\big]\Big) \tag{2.32}$$

$$\begin{aligned}\mu^{(\mathbf{dM})}_{\ell,r,t} = \Bigg[&\frac{1}{\sigma_t^2}\big(\tilde{\mathbf{y}}_{\ell,t} - \widetilde{\mathbf{dm}}_{\ell,\backslash r,t}\mathbf{A}_{\backslash r,t} - \widetilde{\mathbf{m}}_{\ell}\mathbf{a}_{n,t} - \mathbf{x}_{\ell,n,t}\big)\tilde{\mathbf{a}}^{\mathrm{T}}_{r,t} \\ &+ \frac{1}{\psi^2_{\ell,r}}\Big([t<T]dm_{\ell,r,(t+1)} + [t>1]dm_{\ell,r,(t-1)}\Big)\Bigg]\eta^2_{\ell,r,t}\end{aligned} \tag{2.33}$$

where $\widetilde{\mathbf{dm}}_{\ell,\backslash r,t}$ denotes the ℓth row of $\mathbf{dM}_t$ whose rth element has been removed, $\widetilde{\mathbf{m}}_{\ell}$ is the ℓth row of $\mathbf{M}$ and $\mathbf{A}_{\backslash r,t}$ is the matrix $\mathbf{A}_t$ without its rth row. The rows of each variability matrix $\mathbf{dM}_t$ can be sampled in parallel to reduce the computational time required by the sampler.

Sampling the label maps and the outliers (Z and $\underline{\mathbf{X}}$). According to (2.4) and section 2.4.2, the outliers admit the following conditional distributions

$$p(\mathbf{x}_{n,t} \mid \mathbf{y}_{n,t}, \Theta_{\backslash\{z_{n,t},\mathbf{x}_{n,t}\}}) = (1-w_{n,t})\delta(\mathbf{x}_{n,t}) + w_{n,t}\mathcal{N}_{\mathbb{R}_+^L}(\boldsymbol{\mu}^{(\mathbf{X})}_{n,t}, \vartheta_t^2\mathbf{I}_L) \tag{2.34}$$

which are mixtures of a Dirac mass at **0** and of truncated multivariate Gaussian distributions, where

$$w_{n,t} = \frac{\tilde{w}_{n,t}}{\tilde{w}_{n,t} + (1-\omega_{n,t})}, \quad \vartheta_t^2 = \frac{\sigma_t^2 s_t^2}{\sigma_t^2 + s_t^2} \tag{2.35}$$

$$\tilde{w}_{n,t} = \frac{\omega_{n,t}}{(s_t^2)^{L/2}}(\vartheta_t^2)^{L/2}\exp\left(\frac{1}{2\vartheta_t^2}\|\boldsymbol{\mu}^{(\mathbf{X})}_{n,t}\|_2^2\right) \tag{2.36}$$

$$\boldsymbol{\mu}^{(\mathbf{X})}_{n,t} = \frac{s_t^2}{\sigma_t^2 + s_t^2}\big[\mathbf{y}_{n,t} - (\mathbf{M}+\mathbf{dM}_t)\mathbf{a}_{n,t}\big]. \tag{2.37}$$

In practice, the labels $z_{n,t}$ are first sampled according to a Bernoulli distribution to select one of the two models for $\mathbf{x}_{n,t}$, with probability $\mathbb{P}[z_{n,t}=1 \mid \mathbf{y}_{n,t}, \Theta_{\backslash\{z_{n,t},\mathbf{x}_{n,t}\}}] = w_{n,t}$. The labels $\mathbf{z}_{n,t}$ can be sampled in parallel using a checkerboard scheme. Finally, the outliers $\mathbf{x}_{n,t}$ can be sampled in parallel to decrease the computational time.

Sampling the outlier variances ($\mathbf{s}^2$). According to section 2.4.3 and (2.19), the conditional law of s_t^2, $t \in \{1,\dots,T\}$, can be easily identified as the following inverse gamma distribution

$$s_t^2 \mid \Theta_{\backslash\{s_t^2\}} \sim \mathcal{IG}\Big(a_{\mathrm{s}} + \frac{\sharp\{n : z_{n,t}=1\}L}{2}, b_{\mathrm{s}} + \frac{1}{2}\|\mathbf{X}_t\|_{\mathrm{F}}^2\Big). \tag{2.38}$$

Sampling the noise variances ($\boldsymbol{\sigma}^2$). Using section 2.4.3 and (2.19), we obtain for $t \in \{1,\dots,T\}$

$$\sigma_t^2 \mid \mathbf{Y}_t, \Theta_{\backslash\{\sigma_t^2\}} \sim \mathcal{IG}\Big(a_\sigma + \frac{LN}{2}, b_\sigma + \frac{1}{2}\|\mathbf{Y}_t - (\mathbf{M}+\mathbf{dM}_t)\mathbf{A}_t - \mathbf{X}_t\|_{\mathrm{F}}^2\Big). \tag{2.39}$$

Sampling the variability variances ($\mathbf{\Psi}^2$). Similarly, section 2.4.3 and (2.19) lead to

$$\psi_{\ell,r}^2 \mid \Theta_{\backslash \{\psi_{\ell,r}^2\}} \sim \mathcal{IG}\Big(a_\Psi + \frac{T-1}{2}, b_\Psi + \frac{1}{2}\sum_{t=2}^{T}(dm_{\ell,r,t} - dm_{\ell,r,t-1})^2\Big). \tag{2.40}$$

Sampling the granularity parameters ($\boldsymbol{\beta}$). Provided square images are considered, the partition functions $C(\beta_t)$ admit the following closed-form expressions [Ons44; Gio10]

$$\begin{aligned}\widetilde{C}(\beta_t) &= \frac{1}{2}\log(2\sinh\beta_t) + \frac{1}{2N}\sum_{n=1}^{N}\operatorname{acosh}\Delta_n(\beta_t) + \beta_t\\ \Delta_n(\beta_t) &= v(\beta_t) - C_n,\ v(\beta_t) = \frac{\cosh^2\beta_t}{\sinh\beta_t}\\ \widetilde{C}(\beta_t) &= \frac{1}{N}\log C(\beta_t), \quad C_n = \cos\left(\frac{2n-1}{2N}\pi\right).\end{aligned}$$

In this case, the exact partition function can be used to sample the parameters β_t using Metropolis-Hastings steps. In this chapter, new values of the granularity parameters have been proposed by the following Gaussian random walk

$$\beta_t^* = \beta_t^{(q)} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\big(0, \sigma_\beta^2(t)\big) \tag{2.41}$$

where the parameters $\sigma_\beta^2(t)$ are adjusted during the burn-in iterations to yield acceptance rates in the interval $[0.4, 0.6]$. More precisely, the acceptance rates ρ_t are given by

$$\rho_t = \max\left(1, \frac{C(\beta_t^{(q)})}{C(\beta_t^*)}\frac{\mathbb{1}_{[0,2]}(\beta_t^*)}{\mathbb{1}_{[0,2]}(\beta_t^{(q)})}\exp\left[\frac{1}{2}(\beta_t^* - \beta_t^{(q)})\sum_{n=1}^{N}\sum_{k\in\mathcal{V}(n)}\delta(z_{n,t} - z_{k,t})\right]\right).$$

2.5.2. Bayesian inference and parameter estimation

Similarly to [AMH15], the sequence $\{\Theta^{(q)}\}_{q=N_{\text{bi}}+1}^{N_{\text{MC}}}$ generated by the proposed sampler (i.e., after N_{bi} burn-in iterations) is used to approximate the MMSE estimators of the different unknown parameters $\mathbf{M}$, $\mathbf{A}_t$, $\mathbf{dM}_t$ and $\mathbf{X}_t$ by replacing the expectations by empirical averages, i.e.

$$\begin{aligned}\widehat{\mathbf{M}}^{\text{MMSE}} &\simeq \frac{1}{N_{\text{MC}} - N_{\text{bi}}}\sum_{q=N_{\text{bi}}+1}^{N_{\text{MC}}}\mathbf{M}^{(q)}\\ \widehat{\mathbf{A}}_t^{\text{MMSE}} &\simeq \frac{1}{N_{\text{MC}} - N_{\text{bi}}}\sum_{q=N_{\text{bi}}+1}^{N_{\text{MC}}}\mathbf{A}_t^{(q)}\\ \widehat{\mathbf{dM}}_t^{\text{MMSE}} &\simeq \frac{1}{N_{\text{MC}} - N_{\text{bi}}}\sum_{q=N_{\text{bi}}+1}^{N_{\text{MC}}}\mathbf{dM}_t^{(q)}\\ \widehat{\mathbf{X}}_t^{\text{MMSE}} &\simeq \frac{1}{N_{\text{MC}} - N_{\text{bi}}}\sum_{q=N_{\text{bi}}+1}^{N_{\text{MC}}}\mathbf{X}_t^{(q)}\end{aligned}$$

where only the samples associated with the marginal MAP (mMAP) estimator of the label maps have been used for the estimation. Finally, the following mMAP estimator is considered for the latent labels

$$\widehat{z}_{n,t}^{\text{mMAP}} = \underset{z_{n,t} \in \{0,1\}}{\arg\max}\, p\left(z_{n,t} \mid \mathbf{y}_{n,t}, \Theta_{\backslash\{z_{n,t}\}}\right)$$

which is approximated by

$$\widehat{z}_{n,t}^{\text{mMAP}} \simeq \begin{cases} 0, \text{ if } \sharp\{q > N_{\text{bi}} : z_{n,t}^{(q)} = 0\} \leq \frac{N_{\text{MC}} - N_{\text{bi}}}{2} \\ 1, \text{ otherwise.} \end{cases}$$

The main steps of the proposed sampler are summarized in Algo. 4.

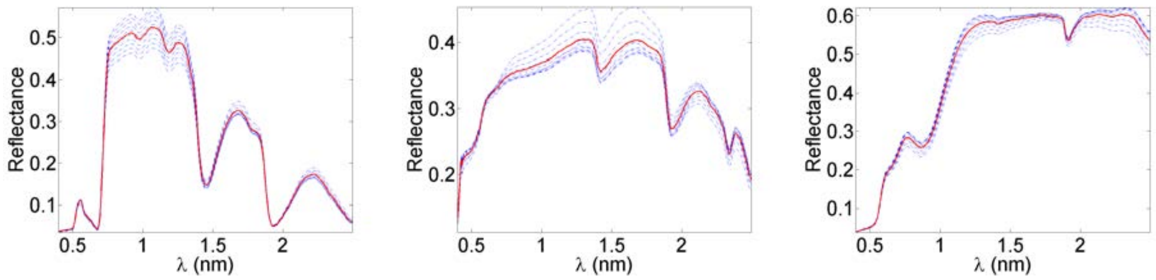
2.5.3. Computational complexity

Assuming elementary scalar arithmetic and pseudo-random number generation are $O(1)$ operations, the overall computational complexity is dominated by matrix products involved in the computation of the variability conditional distribution. Since $R \ll L \ll N$ and $T \ll L$, the per-iteration computational cost of the proposed algorithm is $O(LR^2NT)$. As detailed in the previous paragraphs, many parameters can be sampled in parallel to reduce the computational time of the proposed algorithm. In comparison, the computational complexity of VCA is $O(R^2N)$ [NB05b] per image, and the per iteration complexity of the others algorithms for a single image is respectively: $O(N^2)$ for FCLS [BF10], $O(RN)$ for SISAL [Bio09], $O(LRN)$ for RLMM [FD15] and $O(R^2(L+N))$ for the online unmixing (OU) introduced in the next chapter (see section 3.4.4).

2.6. Experiments with synthetic data

The proposed method has been applied to different MTHS images, for which a new material appears in specific regions of a few images. To this end, less than one half of the images composing the sequence have been corrupted by spatially sparse outliers, which correspond to a new endmember taken from a spectral library. Note that all the images of this experiment do not satisfy the pure pixel assumption to assess the proposed method in challenging situations. A detailed description of the generation procedure is given in the following paragraphs, and the characteristics of each dataset are given in Table 2.1 (in terms of image size, number of endmembers and number of corrupted images).

First scenario. The MTHS images considered in the first scenario reflect the apparition of a new material in half of the images composing the sequences. Each image within a given sequence corresponds to a linear mixture affected by smooth time-varying variability, where the abundances vary smoothly from one image to another. In addition, controlled spectral variability has been introduced by using the product of reference endmembers with randomly generated piecewise-affine functions (as in chapter 1). Different affine functions have been generated for each endmember at each time instant. Typical instances of the signatures used in this experiment are depicted in Fig. 2.2. Finally, the mixtures have been corrupted by an additive white Gaussian noise to ensure a resulting signal-to-noise ratio (SNR) between 25 and 30 dB.



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

Figure 2.2.: Endmembers ($\mathbf{m}_r$, red lines) and their variants affected by variability ($\mathbf{m}_r + \mathbf{dm}_{r,t}$, blue dotted lines) used to generate the synthetic mixtures of $R = 3$ endmembers. Signatures corresponding to different time instants are represented in a single figure to better appreciate the variability introduced in the data.

Second scenario. The second scenario is similar to the first one, except that moderate spatial variability (i.e., occurring within single images) has been additionally introduced in a limited region within each image. In this case, the robustness of the proposed method to moderate spatial variability has been evaluated on two datasets whose characteristics can be found in Table 2.1. To ease the reading of this chapter, these complementary results and the associated comments are deferred to Appendix B.2.

Table 2.1.: Configuration of the synthetic datasets used in the experiments.

	1st scenario			2nd scenario	
Dataset id.	♯1	♯2	♯3	♯4	♯5
R	3	6	9	3	9
T	10	5	5	10	5
Number of corrupted images	4	2	2	4	2
Size (spatial)	50 × 50	100 × 100	100 × 100	50 × 50	100 × 100

2.6.1. Compared methods

The results of the proposed algorithm have been compared to those of several unmixing methods from the literature, some of which are specifically designed to unmix a single HS image. In the following lines, the most relevant implementation details specific to each method are briefly recalled.

1. VCA/FCLS (no variability): the endmembers are first extracted on each image using the vertex component analysis (VCA) [NB05b], which requires pure pixels to be present. The abundances are then estimated for each pixel by solving a fully constrained least squares problem (FCLS) using the Alternating Direction Method of Multipliers (ADMM) [BF10];

2. SISAL/FCLS (no variability): the endmembers are extracted on each image by the simplex identification via split augmented Lagrangian (SISAL) [Bio09], and the abundances are estimated for each pixel by FCLS. The tolerance for the stopping rule has been set to 10^{-3};

3. Robust linear mixing model (RLMM) of [FD15] (no variability): the unmixing method proposed in [FD15] has been independently applied to each image of the series. The algorithm has been initialized with SISAL/FCLS, and the regularization parameter specific to this method is set as suggested in [FD15];

4. Dynamic spectral unmixing (DSU) [HCJ16]: the BCD/ADMM algorithm proposed in [HCJ16], which is based on the ELMM [Dru+16] briefly discussed in the first chapter, has been applied to the entire dataset. The endmembers used to unmix the complete MTHS image are first initialized by VCA applied to the first image of the sequence, and the abundances are initialized with FCLS. To ensure an appropriate comparison with the other methods, the variability extracted at time instant t is interpreted as the deviation of the endmember matrix extracted for the image t from the endmember temporal average;

5. Online unmixing (OU, introduced in the next chapter): the endmembers used to unmix the complete MTHS image are initialized by VCA applied to the first image of the sequence. The abundances are then initialized by FCLS, and the variability matrices are initialized with all their entries equal to 0. The other parameters are set to the same values as those given in the next chapter in Table 3.1;

6. Proposed approach: the endmembers are initialized with VCA applied to the first image of the sequence. The abundances are initialized with FCLS, the variability matrices and label maps are initialized with all their entries equal to 0 (i.e., the images are *a priori* assumed to contain no outlier). The values chosen for the other parameters are summarized in Table 2.2. Further details related to the choice of these values can be found in Appendix B.1;

The performance of the unmixing strategies has been assessed in terms of abundance, endmember and variability estimation errors based on the quality measures given in the preliminary chapter (see section I.4). The results are reported in Table 2.3.

2.6.2. Results

The endmembers estimated by the proposed algorithm are compared to those of VCA/FCLS, SISAL/FCLS, RLMM, DSU and OU in Fig. 2.3, whereas the corresponding abundance maps are displayed in Figs. 2.4 to 2.6. Note that the abundance maps and the endmembers obtained for the dataset $\sharp 4$ (in presence of spatial variability) are deferred to the Appendix B.2. The unmixing performance of each method, reported in Table 2.3 for the first scenario and in Appendix B.2 for the second scenario, leads to the following conclusions.

- **Endmember estimation:** the proposed method shows an interesting robustness with respect to spatially sparse outliers in the sense that the estimated signatures (Figs. 2.3(p) to 2.3(r)) are very close to the corresponding ground truth (Fig. 2.2). In comparison, the shape of the endmembers recovered by VCA, SISAL and RLMM and the variability extracted by the DSU and OU algorithms are significantly affected by outliers, as exemplified in Figs. 2.3(b), 2.3(e), 2.3(h), 2.3(k) and 2.3(n) respectively. These qualitative results are confirmed by the quantitative performance measures of each method provided in Table 2.3. Note that the endmembers recovered by the SISAL and RLMM methods are very sensitive to the VCA initialization, as illustrated by the similarity between the signatures estimated by these methods (Figs. 2.3(a) to 2.3(i)).

- **Abundance estimation:** the abundance maps estimated by FCLS, RLMM and SISAL reflect the high sensitivity of VCA (used to initialize SISAL and RLMM) to the presence of outliers (see the figures delineated in red in Fig. 2.4). On the contrary, the abundances recovered by OU and the proposed approach are much closer to the ground truth. These observations are confirmed by the abundance estimation performance reported in Table 2.3.
- **Overall performance:** the performance measures reported in Table 2.3 are globally favorable to the proposed approach. It is important to mention that the price to pay with the good performance of the proposed method is its computational complexity, which is common with MCMC methods.

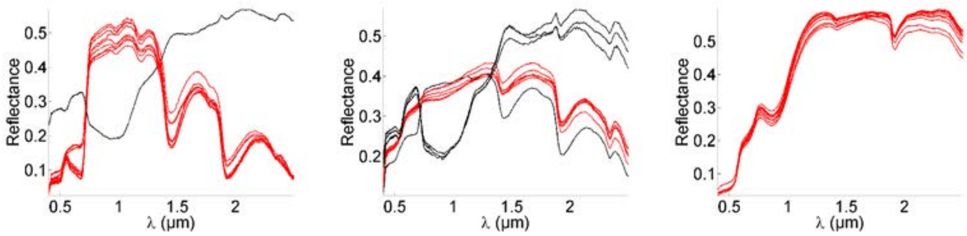
As a complementary output, the proposed algorithm is able to recover the location of the outliers within each image, as illustrated in Fig. 2.7. Up to a few false alarms, the estimated labels are very close to the ground truth.

Table 2.2.: Fixed parameters, and initial values associated in the experiments to parameters later inferred from the model.

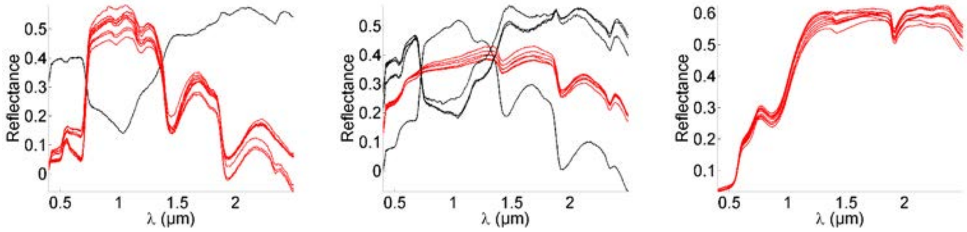
	Parameters	Synthetic data	Real data
Fixed parameters	ε_n^2	10^{-3}	10^{-2}
	ξ	1	1
	ν	10^{-3}	10^{-5}
	$a_{\mathbf{s}}, a_{\Psi}, a_{\sigma}$	10^{-3}	10^{-3}
	$b_{\mathbf{s}}, b_{\Psi}, b_{\sigma}$	10^{-3}	10^{-3}
	N_{bi}	350	450
	N_{MC}	400	500
Initial values	σ_t^2	10^{-4}	10^{-4}
	s_t^2	5×10^{-3}	5×10^{-3}
	$\psi_{\ell,r}^2$	10^{-3}	10^{-2}
	β_t	1.7	1.7

Table 2.3.: Simulation results on synthetic multi-temporal data (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$).

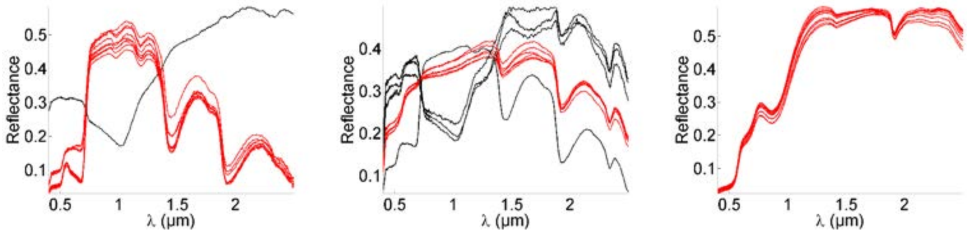
		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	time (s)
$R = 3$ (♯1)	VCA/FCLS	6.07	2.32	/	3.91	**1**
	SISAL/FCLS	5.07	1.71	/	2.28	2
	RLMM	5.13	2.04	/	**0.31**	463
	DSU	5.18	0.53	11.5	2.21	8
	OU	**1.90**	0.42	3.22	2.61	98
	Proposed	2.03	**0.15**	**1.85**	2.00	2530
$R = 6$ (♯2)	VCA/FCLS	3.81	1.57	/	3.09	**2**
	SISAL/FCLS	5.76	0.91	/	4.49	3
	RLMM	2.73	1.26	/	**0.29**	1453
	DSU	4.54	1.27	4.34	0.62	22
	OU	2.74	0.38	3.70	1.13	420
	Proposed	**1.48**	**0.16**	**2.84**	0.51	8691
$R = 9$ (♯3)	VCA/FCLS	3.74	0.65	/	6.83	**4**
	SISAL/FCLS	5.91	0.36	/	5.56	5
	RLMM	2.48	0.54	/	**0.31**	1447
	DSU	**1.96**	0.18	2.20	0.85	12
	OU	6.08	0.47	**2.19**	0.89	1024
	Proposed	2.23	**0.15**	8.38	0.82	17151



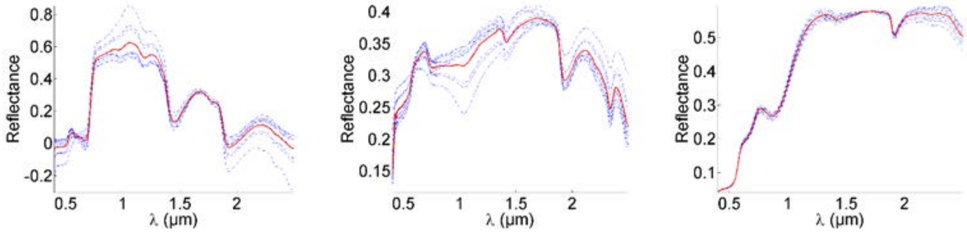
(a) VCA (b) VCA (c) VCA



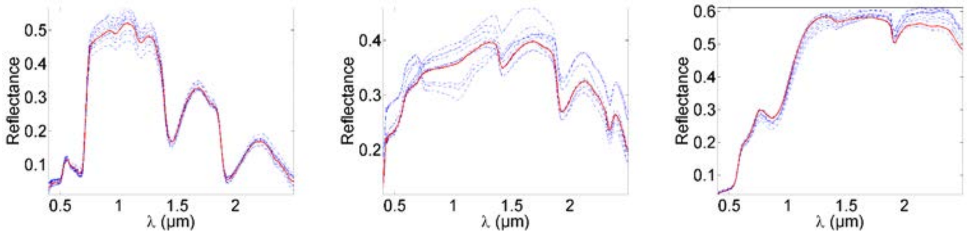
(d) SISAL (e) SISAL (f) SISAL



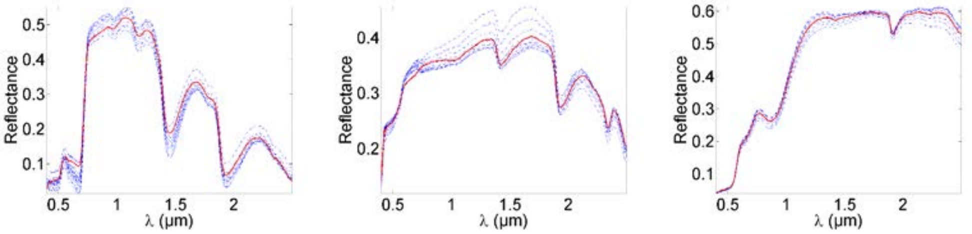
(g) RLMM (h) RLMM (i) RLMM



(j) DSU (k) DSU (l) DSU



(m) OU (n) OU (o) OU



(p) MCMC (q) MCMC (r) MCMC

Figure 2.3.: Endmembers ($\mathbf{m}_r$, red lines) and their variants affected by variability ($\mathbf{m}_r + \mathbf{dm}_{r,t}$, blue dotted lines) recovered by the different methods from the synthetic dataset $\sharp 1$. Signatures corresponding to different time instants are represented on a single figure to better appreciate the variability recovered from the data. The spectra represented in black correspond to signatures significantly affected by outliers.

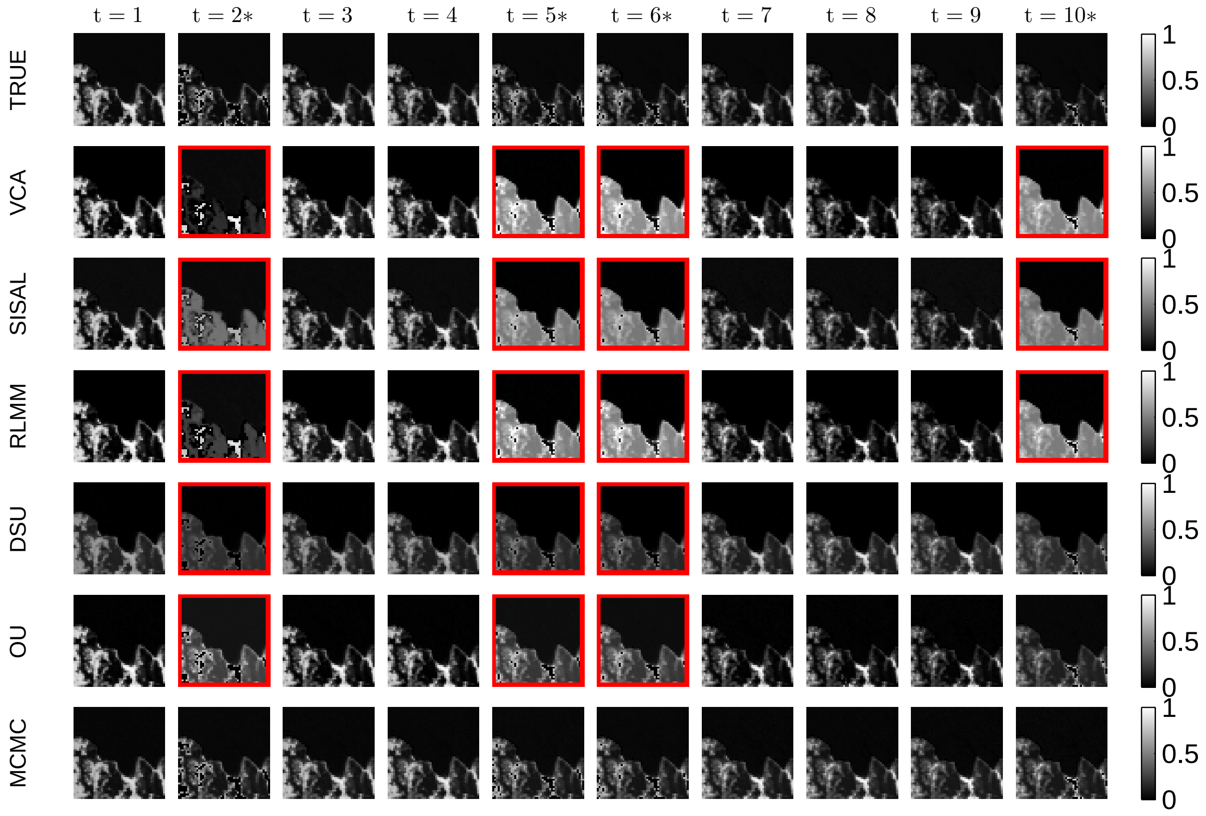


Figure 2.4.: Abundance map of the first endmember recovered by the different methods (in each row) at each time instant (given in column) for the experiment ♯1 [the different rows correspond to the true abundances, VCA/FCLS, SISAL/FCLS, RLMM, DSU, OU and the proposed method]. The images delineated in red show that several methods are highly sensitive to the presence of outliers, and the time instants represented with ∗ indicate images containing outliers.

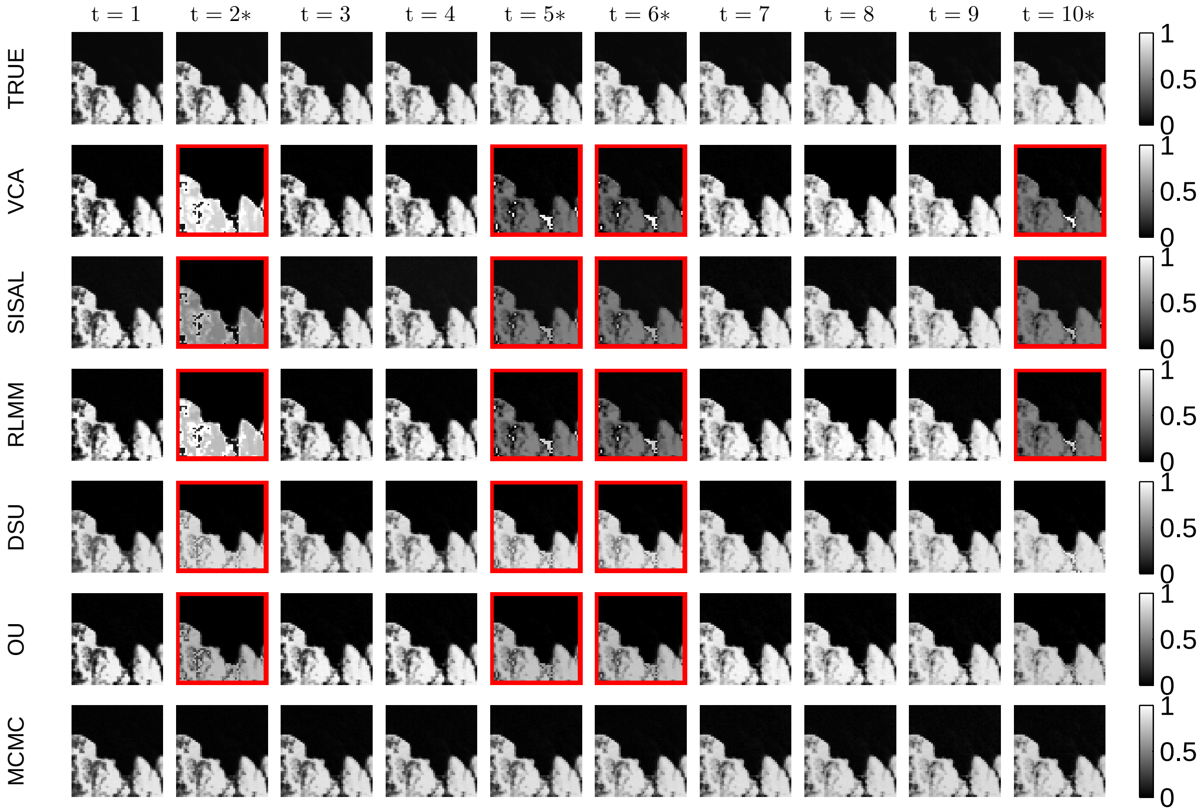


Figure 2.5.: Abundance map of the second endmember recovered by the different methods at each time instant for the experiment ♯1. The images delineated in red show that several methods are highly sensitive to the presence of outliers, and the time instants represented with $*$ indicate images containing outliers.

Figure 2.6.: Abundance map of the third endmember recovered by the different methods at each time instant for the experiment ♯1. The images delineated in red show that several methods are highly sensitive to the presence of outliers, and the time instants represented with $*$ indicate images containing outliers.

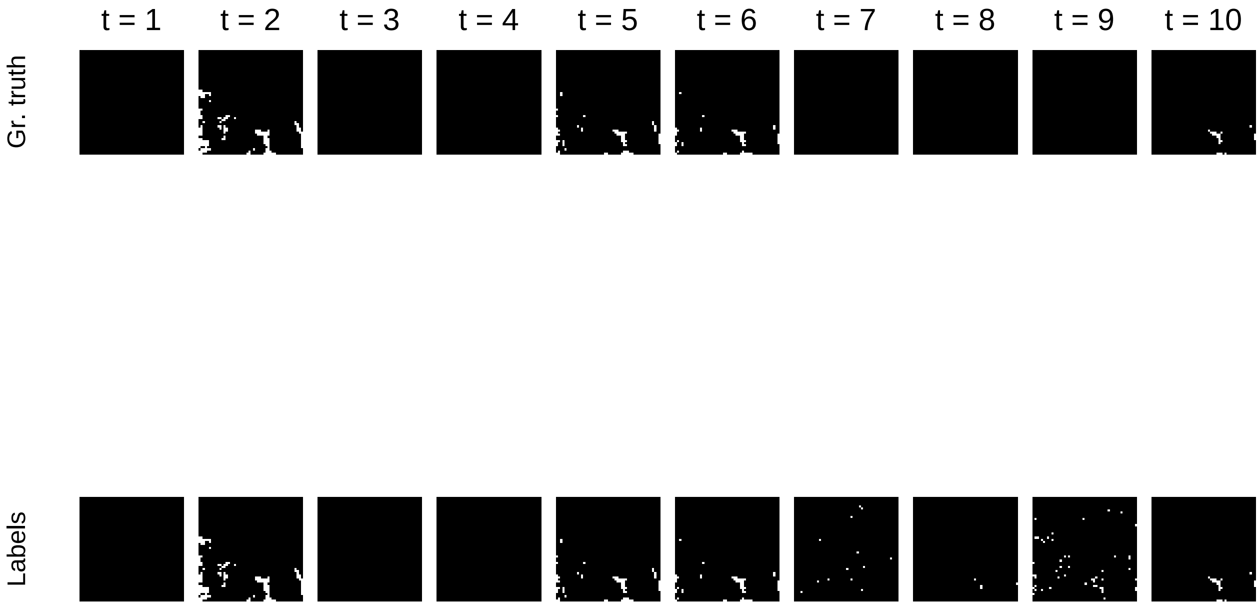


Figure 2.7.: Outlier labels $\mathbf{z}_t$ estimated for each image of the synthetic dataset with 3 endmembers [0 in black, 1 in white].

2.7. Experiments with real data

2.7.1. Description of the dataset

We consider a real sequence of HS images acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over the Lake Tahoe region (Mud Lake, California, United States of America) between 2014 and 2015[1]. The scene of interest (100×100), composed of a lake and a nearby field, has been unmixed with $R = 3$ endmembers based on the results of the noise-whitened eigengap algorithm (NWEGA) [Hal+16b] applied to each image of the series (see Table 2.4). For $R = 4$ and $R = 5$, the signatures of water, soil and vegetation were split into two or more components by the different algorithms, suggesting $R = 3$ is more appropriate for this study. A careful inspection of the data reveals that this dataset contains outliers (area delineated in red in Fig. 2.8(e)). After removing the seemingly corrupted bands and the water absorption bands, 173 out of the 224 spectral bands were finally exploited. The initial parameters used for the proposed algorithm are given in Table 2.2, while the other methods have been run with the same parameters as in section 2.6.


(a) 10/04/2014 (b) 02/06/2014 (c) 19/09/2014 (d) 17/11/2014 (e) 29/04/2015 (f) 16/10/2015

Figure 2.8.: Scenes used in the experiment, given with their respective acquisition date. The area delineated in red in Fig. 2.8(e) highlights a region known to contain outliers.

Table 2.4.: Endmember number R estimated by NWEGA [Hal+16b] on each image of the real dataset.

	04/10/2014	06/02/2014	09/19/2014	11/17/2014	04/29/2015	10/13/2015
NWEGA	3	3	3	4	3	4

2.7.2. Results

Given the absence of ground truth, the performance of the unmixing methods is assessed in terms of RE (Table 2.5) while taking into account the consistency of the estimated abundance maps reported in Figs. 2.9, 2.10 and 2.11. More precisely, the abundances associated with the vegetation area are expected to be very high for $t = 1, 3, 5$ (corresponding to Figs. 2.8(a), 2.8(c) and 2.8(e)) where the vegetation visually appears to be sufficiently irrigated (hence well represented). On the contrary, the abundance coefficients are supposed to be much lower for $t = 2, 4, 6$ (corresponding to Figs. 2.8(b), 2.8(d) and 2.8(f)), where the vegetation is visually drier or almost absent. Concerning the presence of water in the bottom

---

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

left-hand corner of the images, the latent variables introduced in section 2.4.2 are expected to reflect the abrupt variations in the presence of water observed at $t = 3$, 4 and 5. These observations, combined with the extracted signatures (Fig. 2.12) and the estimated abundances (Figs. 2.9 to 2.11) lead to the following comments.

- **Endmember estimation:** the signature recovered for the soil by VCA, SISAL and RLMM at time $t = 5$ shows an amplitude which is significantly larger than the other signatures, and a shape incompatible with what can be expected based on physical considerations (see the black signatures in Figs. 2.12(a), 2.12(d) and 2.12(g)). This is a clear indication that outliers are present in the corresponding image. A similar observation can be made for the vegetation signature obtained by VCA, SISAL and RLMM at time $t = 5$. On the contrary, the endmembers recovered by DSU, OU and the proposed approach are much more consistent from this point of view.

- **Abundance estimation:** the estimated abundances globally reflect the previous comments made on the extracted endmembers. Notably, the abundance coefficients estimated at $t = 5$ by VCA, SISAL and RLMM (delineated in red in Figs. 2.9 to 2.11) are visually inconsistent with the temporal evolution of the materials observed in the true color composition given in Fig. 2.8. More explicitly, the soil is not supposed to be concentrated on a few pixels as suggested by the corresponding abundance maps in Fig. 2.9. Similarly, the water is not supposed to be present in high proportions in all the pixels of the image as indicated in Fig. 2.10. These results, in contradiction with Fig. 2.8, suggest that outliers are present at $t = 5$. In addition, the abundance maps estimated at $t = 4$ and 6 by FCLS for the soil and the water (delineated in green in Figs. 2.9 and 2.10) suggest that the water contribution has been split into two spectra. The corresponding signatures are represented in green in Figs. 2.12(a) and 2.12(c). On the contrary, the results reported for DSU, OU and the proposed method are consistent with the expected evolution of water and vegetation over time (abundance values close to 1 at time $t = 1, 3, 5$, lower values at time $t = 2, 4, 6$). Finally, the vegetation abundance maps estimated by the proposed method globally presents a better contrast than those obtained with DSU and OU (Fig. 2.11).

Table 2.5.: Simulation results on real data (RE $\times 10^{-4}$).

		RE	time (s)
$R = 3$	VCA/FCLS	45.05	**1**
	SISAL/FCLS	1.65	2
	RLMM	2.51	390
	DSU	6.03	26
	OU	2.50	508
	Proposed	**0.34**	23608

The previous comments, along with the lower reconstruction error reported in Table 2.5, suggest that the proposed approach is robust to spatially sparse outliers while allowing smooth temporal variations to be exploited. Indeed, the pixels corresponding to abrupt variations of the water signature have been properly detected. Furthermore, the outliers detected in this dataset for $t = 5$ (highlighted in red in Fig. 2.8(e)) are well captured by the latent variables $\mathbf{Z}$ (see Fig. 2.13). In addition, the spatial distribution of the estimated outlier labels (Fig. 2.13) is in agreement with the results of

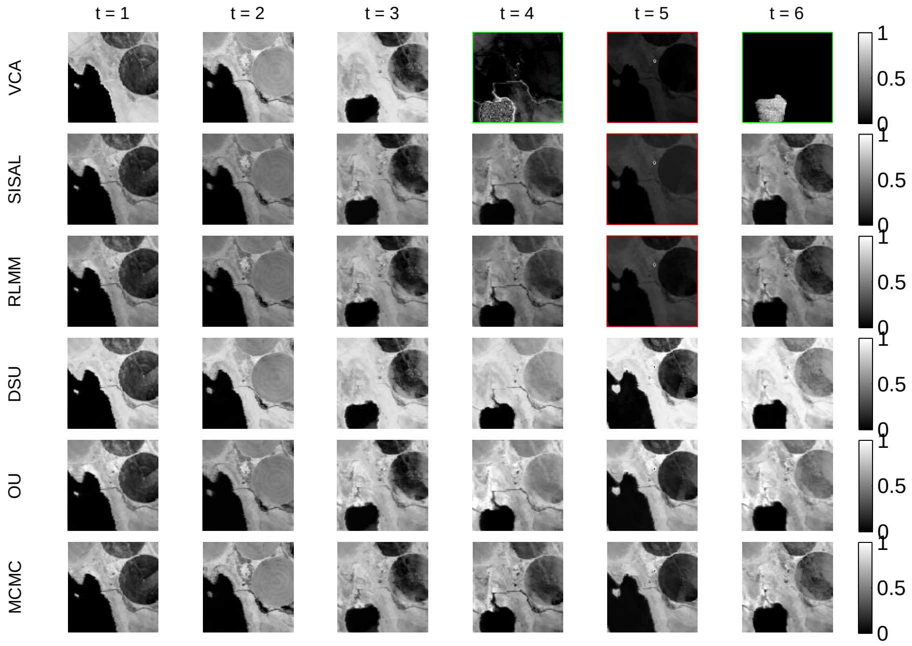


Figure 2.9.: Soil abundance map recovered by the different methods (in each row) at each time instant (given in column) for the experiment on the real dataset. On the one hand, the images delineated in red suggest that some of the methods are particularly sensitive to the presence of outliers. On the other hand, the images delineated in green represent the abundance maps associated with signatures which have been split into two components by the corresponding unmixing procedures.

the RLMM (in terms of the spatial distribution of the outlier energy) and with the non-linearity detector [Alt+13] applied to each image of the sequence with the SISAL-estimated endmembers (see Fig. 2.14). Concentrated on regions where non-linear effects can be reasonably expected, the active latent variables $\mathbf{Z}$ tend to capture the spatial distribution of the non-linearities possibly occurring in the observed scene.

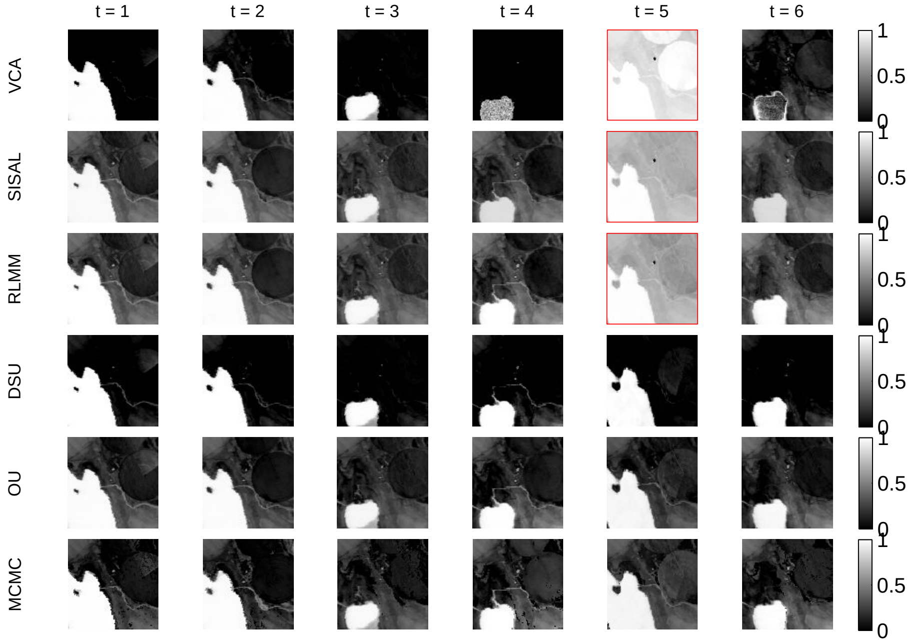


Figure 2.10.: Water abundance map recovered by the different methods (in each row) at each time instant (given in column) for the experiment on the real dataset. The images delineated in red suggest that some of the methods are particularly sensitive to the presence of outliers.

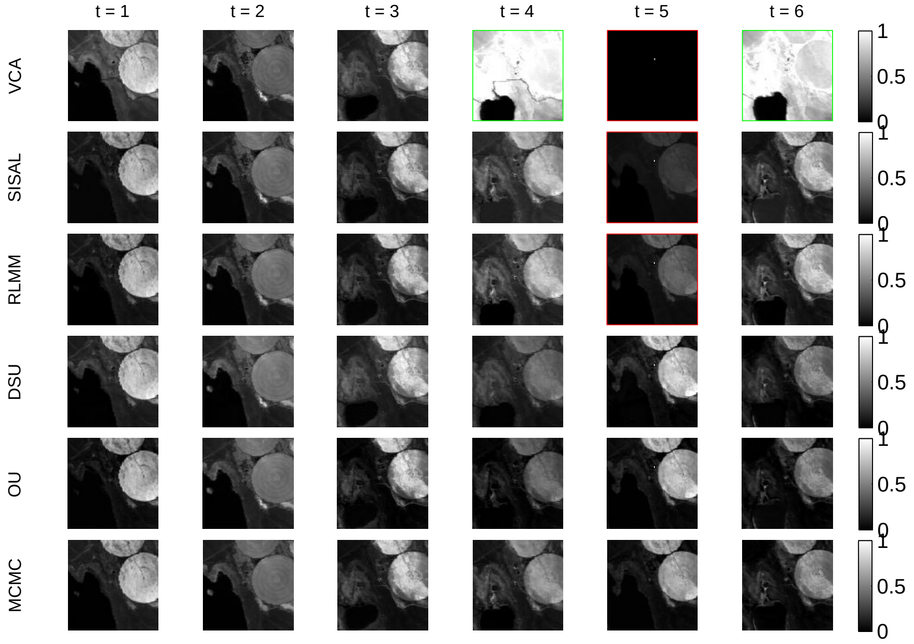
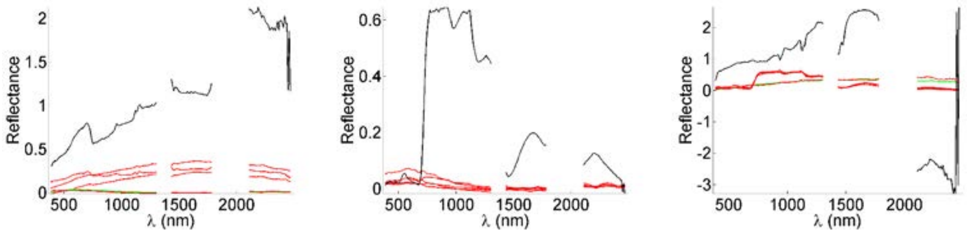
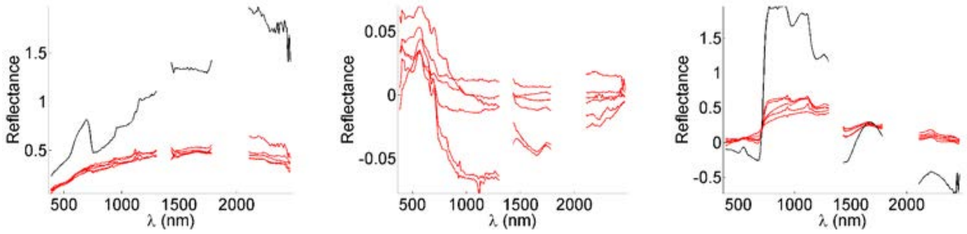


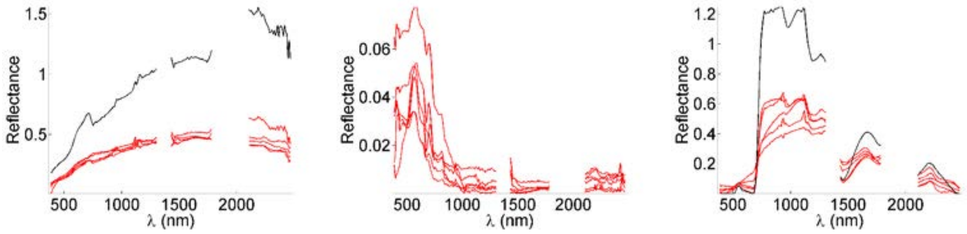
Figure 2.11.: Vegetation abundance map recovered by the different methods at each time instant for the experiment on the real dataset. On the one hand, the images delineated in red suggest that some of the methods are particularly sensitive to the presence of outliers. On the other hand, the images delineated in green represent the abundance maps associated with signatures which have been split into two components by the corresponding unmixing procedures.



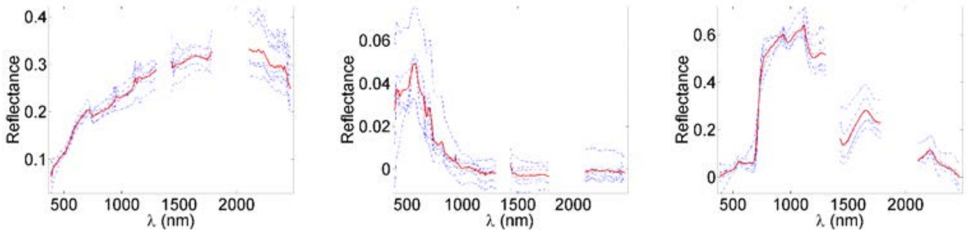
(a) VCA (b) VCA (c) VCA



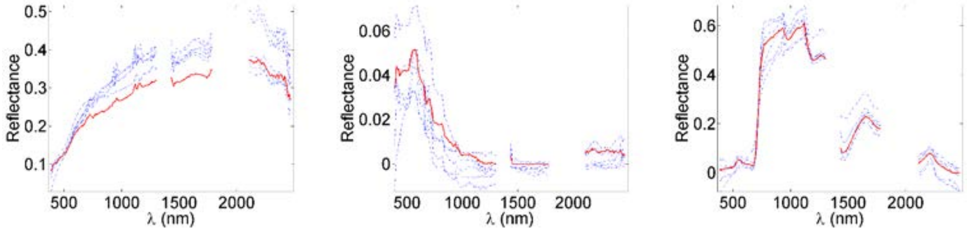
(d) SISAL (e) SISAL (f) SISAL



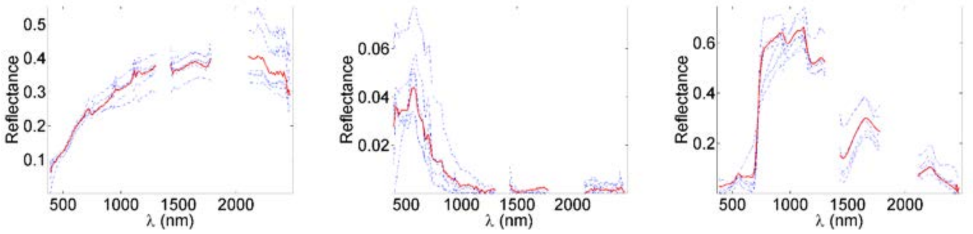
(g) RLMM (h) RLMM (i) RLMM



(j) DSU (k) DSU (l) DSU



(m) OU (n) OU (o) OU



(p) MCMC (q) MCMC (r) MCMC

Figure 2.12.: Endmembers ($\mathbf{m}_r$, red lines) and their variants affected by variability ($\mathbf{m}_r + \mathbf{dm}_{r,t}$, blue dotted lines) recovered by the different methods from the real dataset depicted in Fig. 2.8. Signatures corresponding to different time instants are represented in a single figure to better appreciate the variability recovered from the data. The spectra represented in black correspond to signatures corrupted by outliers, while those given in green represent endmembers which have been split into several components by the associated estimation procedure.

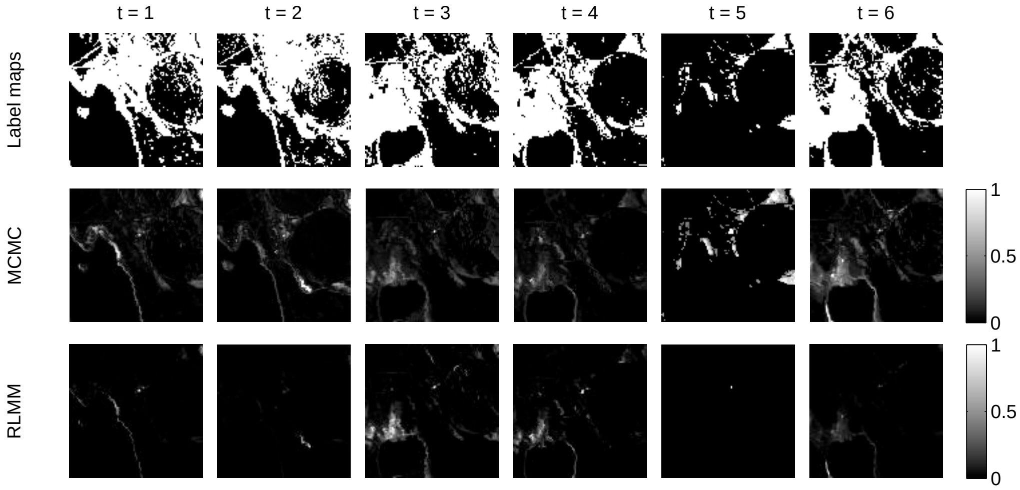


Figure 2.13.: mMAP estimates of the label maps recovered by the proposed approach, displayed at each time instant (the different rows correspond to: the estimated label map (pixels detected as outliers appear in white), the outlier energy map re-scaled in the interval $[0, 1]$ obtained by the proposed method, and by RLMM).

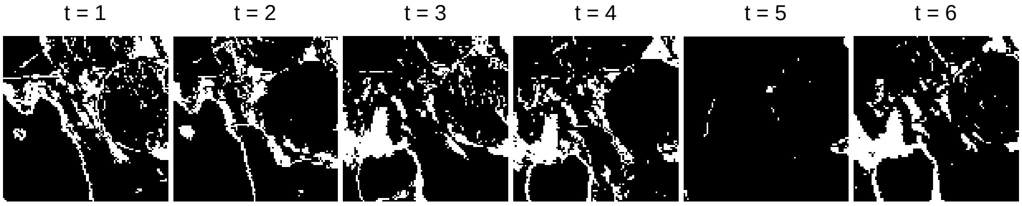


Figure 2.14.: Non-linearity maps estimated by the detector [Alt+13] applied to each image with the SISAL-extracted endmembers, with a probability of false alarm of 10^{-3} (pixels detected as non-linearities appear in white).

2.8. Conclusion and future work

This chapter introduced a new Bayesian model accounting for both smooth and abrupt variations possibly occurring in MTHS images. In particular, the adopted model was designed to handle datasets in which mostly the same materials are expected to be observed at different time instants, thus allowing information redundancy to be exploited. An MCMC algorithm was investigated to solve the resulting unmixing problem in order to precisely assess the performance of the proposed approach on MTHS images of moderate size (i.e., moderate spatial and temporal dimensions). This algorithm was used to sample the posterior of the proposed hierarchical Bayesian model, thus allowing estimators of the unknown model parameters to be built. Future research perspectives include the use of relaxation methods to the Ising field [MSB03; BCK14; BF16] to tackle similar problems with online optimization techniques.

However, a major issue results from the possibly significant dimension of MTHS data. In this context, the next chapter is devoted to the design of an online unmixing algorithm to efficiently address endmember temporal variability.

Main contributions. A hierarchical Bayesian model was introduced to analyze multitemporal hyperspectral images. This model accounts for smooth temporal variations – construed as spectral variability – and abrupt spectral changes interpreted as outliers. The parameters of this model are inferred using a Markov chain Monte-Carlo (MCMC) method, which allows the posterior of interest to be sampled and Bayesian estimators to be approximated. The encouraging results obtained on real data lead to the conclusion that MTHS images can provide improved unmixing performance when compared to methods analyzing the images independently.

2.9. Conclusion (in French)

Ce chapitre a étudié un modèle bayésien hiérarchique pour prendre en compte des variations spectrales aussi bien modérées qu'abruptes lors du démélange d'image hyperspectrales multi-temporelles. Le modèle proposé est plus particulièrement adapté à l'analyse de données pour lesquelles les mêmes matériaux sont observés de façon récurrente. Un algorithme de Monte-Carlo par chaîne de Markov a permis de résoudre le problème de démélange issu de ce modèle, conduisant à des performances satisfaisantes ausis bien sur données synthétiques que réelles. En particulier, cette approche a permis d'obtenir une estimation cohérente du support des données aberrantes sur données réelles, en lien avec la connaissance disponible *a priori* sur ces données.

Au vue de la dimension importante des données hyperspectrales multi-temporelles, le développement d'une procédure de démélange en ligne fait l'objet du chapitre suivant. Une perspective de recherche à court terme consisterait à relâcher l'utilisation du champ d'Ising comme proposé dans [MSB03; BCK14; BF16] pour pouvoir appliquer une procédure d'estimation en ligne similaire à celle décrite au chapitre suivant.

Contributions principales. Un modèle bayésien hiérarchique est proposé pour le démélange robuste d'images hyperspectrales multi-temporelles. Ce modèle de mélange rend compte d'une variation temporelle modérée des spectres purs

– décrite en termes de variabilité – et de changements spectraux abrupts – interprétés en termes de données aberrantes. Le modèle proposé s'avère particulièrement pertinent pour l'analyse de données réelles, notamment lorsque l'intervalle de temps entre deux acquisitions est important. Dans un tel cas, l'apparition d'un matériau non présent dans la scène de référence induit par exemple des variations spectrales abruptes pouvant être interprétées comme des données aberrantes.

CHAPTER 3

Online unmixing of multitemporal hyperspectral images

This chapter has been adapted from the journal paper [TDT16b].

Contents

3.1. Introduction (in French)

D’après l’étude menée au chapitre précédent, les images hyperspectrales multi-temporelles – composées de plusieurs images hyperspectrales acquises pour une même région à différents instants – présentent un intérêt notable dans la mesure où elles permettent d’exploiter une certaine redondance de l’information apportée par des images consécutives (grâce aux variations temporelles modérées de certains des paramètres du mélange [HMT15; Hal+16a]). En particulier, le précédent chapitre a illustré la pertinence des données multi-temporelles lorsque les spectres purs sont correctement représentés dans différentes images. Par ailleurs, elles ont permis d’obtenir une caractérisation satisfaisante de la variabilité temporelle.

Toutefois, l'utilisation de données multi-temporelles peut compromettre l'utilisation de procédures de démélange nécessitant d'exploiter directement l'ensemble des données [Hal+15; HCJ16; YZP17], notamment en raison de contraintes matérielles (e.g., ressources calculatoires limitées). Ce chapitre est ainsi consacré à la mise en œuvre d'un algorithme de démélange en ligne, i.e., qui permet d'incorporer progressivement les données disponibles de façon séquentielle dans la procédure de démélange. Comme les spectres purs extraits à une date donnée peuvent être *in fine* considérés comme la réalisation de spectres de référence, une variante du modèle introduit dans le chapitre 1 est utilisée pour prendre en compte la variabilité temporelle des sources. Inspiré par les travaux [RX11; Mai+10], la formulation du problème est proposée sous la forme d'un problème d'optimisation stochastique avec recours. Une deuxième interprétation de l'algorithme proposé est également envisagée dans le cadre des travaux présentés dans [CPR16].

Ce chapitre s'organise de la façon suivante. Le modèle considéré dans ce chapitre est en premier lieu décrit dans la section 3.3. La section 3.4 se concentre sur la formulation du problème, ainsi que la description de l'algorithme en ligne envisagé. Une interprétation du problème dans le cadre de l'algorithme d'optimisation introduit dans [CPR16] y est par ailleurs proposée. Les résultats numériques obtenus sur données synthétiques et réelles sont respectivement rapportés dans les sections 3.5 et 3.6. À ce titre, les résultats sont comparés à ceux de quatre méthodes de la littérature appliquées à chacune des images de façon indépendante (VCA [NB05b] / FCLS [HC01; BF10]), SISAL [Bio09] / FCLS, $\ell_{1/2}$-NMF [Qia+11] et l'algorithme BCD/ADMM introduit au chapitre 1), ainsi qu'à l'algorithme DSU [HCJ16] et PALM [BST13] appliqués (dès que possible) à l'ensemble des données disponibles. La section 3.7 vient finalement conclure l'étude menée dans ce chapitre.

3.2. Introduction

Following the study conducted in the previous chapter, sequences of HS images acquired over the same area at different time instants can be of interest to exploit information redundancy between consecutive images (e.g., through features exhibiting moderate or smooth temporal variations [HMT15; Hal+16a]). More precisely, chapter 2 has shown that the use of several images is particularly appealing to improve the unmixing results when all the observed materials are not well represented in each image. The results obtained on real datasets outperformed those relying on an individual analysis of each image [Goe+13], while allowing the endmembers temporal evolution to be characterized.

However, the use of MTHS data may preclude the application of batch unmixing techniques, such as the method considered in the preceding chapter or in [Hal+15; HCJ16; YZP17]. Indeed, the large size of these data can lead to several issues, related to limits in the available memory or computational resources. This chapter focuses on the design of an online unmixing algorithm, which allows data to be sequentially incorporated into the estimation process without the need to simultaneously load all the data into memory. Since the identified endmembers can be considered as time-varying instances of reference endmembers, we consider the perturbed linear mixing model (PLMM) introduced in chapter 1 to account for spectral variability. However, inspired by the works presented in [RX11; Mai+10], the unmixing problem is formulated as a two-stage stochastic program. A different formulation of the problem, based on the optimization framework considered in [CPR16], is also investigated.

The present chapter is organized as follows. The proposed PLMM accounting for temporal variability is introduced in section 3.3. Section 3.4 describes an online algorithm to solve the resulting optimization problem. Another formulation of the problem, considered as an application of [CPR16], is also investigated. Experimental results obtained on synthetic and real data are reported in sections 3.5 and 3.6 respectively. The results obtained with the proposed algorithm are systematically compared to those obtained with VCA [NB05b] / FCLS [HC01; BF10]), SISAL [Bio09] / FCLS, the $\ell_{1/2}$ non-negative matrix factorization (NMF) [Qia+11] and the BCD/ADMM algorithm introduced in chapter 1, each method being independently applied to each image of the sequence. A comparison with the DSU [HCJ16] approach and a batch PALM [BST13] algorithm is also proposed whenever possible. Section 3.7 provides preliminary conclusions on this chapter, and highlights future research perspectives.

3.3. Problem statement

3.3.1. Perturbed linear mixing model (PLMM)

Throughout this chapter, we consider a multitemporal hyperspectral (MTHS) image, composed of T HS images acquired over the same area at different time instants. In the following, at most R endmembers are assumed to be present in – and common to – the different images. In fact, each endmember does not need to be present in each image, but at least in one image of the time series. In practice, the unmixing results first obtained from a reference image are used as a starting point to unmix the whole sequence. The proposed model addresses the case where the variability essentially results from the global acquisition conditions or from the evolution of the scene between consecutive images. As a first approximation, the variability is thus assumed to be constant within each image as in the previous chapter. The resulting model can thus be written

$$\mathbf{y}_{n,t} = \sum_{r=1}^{R} a_{r,n,t}\big(\mathbf{m}_r + \mathbf{dm}_{r,t}\big) + \mathbf{b}_{n,t} \tag{3.1}$$

for $n = 1, \ldots, N$ and $t = 1, \ldots, T$, where $\mathbf{y}_{n,t}$ denotes the nth image pixel at time t, $\mathbf{m}_r$ is the rth endmember, $a_{r,n,t}$ is the proportion of the rth endmember in the nth pixel at time t, and $\mathbf{dm}_{r,t}$ denotes the perturbation of the rth endmember at time t. Finally, $\mathbf{b}_{n,t}$ models the noise resulting from the data acquisition and the modeling errors. In matrix form, the model (3.1) can be expressed as

$$\mathbf{Y}_t = (\mathbf{M} + \mathbf{dM}_t)\mathbf{A}_t + \mathbf{B}_t \tag{3.2}$$

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

$$\mathbf{M} \succeq \mathbf{0}_{L,R}, \quad \mathbf{A}_t \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}_t^{\mathrm{T}}\mathbf{1}_R = \mathbf{1}_N, \forall t = 1, \dots, T \tag{3.3}$$

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

$$\|\mathbf{dM}_t\|_{\mathrm{F}} \leq \nu, \quad \text{for } t = 1, \dots, T \tag{3.4}$$

$$\frac{1}{T}\left\|\sum_{t=1}^{T}\mathbf{dM}_t\right\|_{\mathrm{F}} \leq \kappa \tag{3.5}$$

where ν and κ are fixed positive constants, and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. To highlight their practical interest, the two preceding constraints can be interpreted in terms of the feasible domain of $\mathbf{M}$ and $\mathbf{dM}_t$. Introducing the perturbed endmembers $\mathbf{M}_t \triangleq \mathbf{M} + \mathbf{dM}_t$, the constraint (3.4) can be reformulated as

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

where $\mathcal{B}_{\mathrm{F}}(\mathbf{M}_t, \nu)$ is the ball of center $\mathbf{M}_t$ and of radius ν. This reformulation shows 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, (3.5) constrains the perturbed endmembers to be distributed around the true endmembers (since the average of the different variabilities is small), i.e., the endmember signatures $\mathbf{M}$ should reflect the average behavior of the perturbed endmembers $\mathbf{M}_t$. In practice, setting ν to a reasonable value is desirable from a modeling point of view. Indeed, large perturbations are usually interpreted as outliers that can be removed prior to the unmixing process, or which should be specifically captured as in chapter 2. Note however that the algorithm proposed in section 3.4.2 is independent from any consideration on the values of ν and κ.

Remark 3.1. Since HS unmixing is performed on reflectance data, $\mathbf{Y}_t \in [0,1]^{L\times N}$, and 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 to the convergence result given in paragraph 3.4.3. 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.

3.3.2. Problem formulation

To estimate the endmembers composing the image sequence, the model (3.1) and the constraints (3.3)–(3.5) can be combined to formulate a two-stage stochastic program, leading to the design of an online estimation algorithm. 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 function only depends on the endmembers. Assuming the expectations are well-defined, we consider the following optimization problem

$$\min_{\mathbf{M}\in\mathcal{M}} g(\mathbf{M}) = \mathbb{E}_{\mathbf{Y},\mathbf{A},\mathbf{dM}}\big[f(\mathbf{Y},\mathbf{M},\mathbf{A},\mathbf{dM})\big],\ \mathcal{M}=[0,1]^{L\times R}. \tag{3.6}$$

The function f is defined as

$$f(\mathbf{Y},\mathbf{M},\mathbf{A},\mathbf{dM}) = \frac{1}{2}\left\|\mathbf{Y}-(\mathbf{M}+\mathbf{dM})\mathbf{A}\right\|_{\mathrm{F}}^2 + \alpha\Phi(\mathbf{A}) + \beta\Psi(\mathbf{M}) + \gamma\Upsilon(\mathbf{dM}) \tag{3.7}$$

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

$$\mathbf{A}\in\mathcal{A}_R = \left\{\mathbf{A}\in\mathbb{R}^{R\times N}\,\middle|\,\mathbf{a}_n\in\mathcal{S}_R,\ n=1,\dots,N\right\} \tag{3.8}$$

$$\mathbf{dM}\in\mathcal{D} = \mathcal{B}_{\mathrm{F}}(\mathbf{0},\nu)\cap\left\{\mathbf{dM}\mid\left\|\mathbb{E}[\mathbf{dM}]\right\|_{\mathrm{F}}\leq\kappa\right\}. \tag{3.9}$$

The parameters α, β and γ ensure a trade-off between the data fitting term and the penalization functions Φ, Ψ, Υ. In practice, g is approximated at time t by an upper bound $\hat{g}_t$ given by a stochastic approximation [Mai+10]

$$\begin{aligned}\hat{g}_t(\mathbf{M}) &= \frac{1}{2t}\sum_{i=1}^{t}\left\|\mathbf{Y}_i-(\mathbf{M}+\mathbf{dM}_i)\mathbf{A}_i\right\|_{\mathrm{F}}^2 + \beta\Psi(\mathbf{M})\\ &= \frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{2}\left\|\mathbf{M}\mathbf{A}_i\right\|_{\mathrm{F}}^2 - \langle\mathbf{Y}_i-\mathbf{dM}_i\mathbf{A}_i,\mathbf{M}\mathbf{A}_i\rangle\right) + \beta\Psi(\mathbf{M}) + c\\ &= \frac{1}{t}\left[\frac{1}{2}\mathrm{Tr}(\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{C}_t) + \mathrm{Tr}(\mathbf{M}^{\mathrm{T}}\mathbf{D}_t)\right] + \beta\Psi(\mathbf{M}) + c \end{aligned} \tag{3.10}$$

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

$$\mathbf{C}_t = \sum_{i=1}^{t}\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}},\quad \mathbf{D}_t = \sum_{i=1}^{t}(\mathbf{dM}_i\mathbf{A}_i-\mathbf{Y}_i)\mathbf{A}_i^{\mathrm{T}}. \tag{3.11}$$

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

$$\mathcal{D}_t = \mathcal{B}_{\mathrm{F}}(\mathbf{0},\nu)\cap\left\{\mathbf{dM}\mid\left\|\mathbf{dM}+\mathbf{E}_{t-1}\right\|_{\mathrm{F}}\leq t\kappa\right\} \tag{3.12}$$

with

$$\mathbf{E}_t = \sum_{i=1}^{t}\mathbf{dM}_i. \tag{3.13}$$

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

Abundance penalization. The abundance penalization Φ considered in this chapter has been chosen to promote temporally smooth abundances – in the ℓ_2-norm sense – between two consecutive images

$$\Phi(\mathbf{A}_t) = \frac{1}{2}\left\|\mathbf{A}_t-\mathbf{A}_{t-1}\right\|_{\mathrm{F}}^2. \tag{3.14}$$

As long as Φ satisfies the regularity condition given in paragraph 3.4.3, any other type of prior knowledge relative to the abundances can be incorporated into the proposed method.

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 chapter, we consider the mutual distance between each endmember introduced in [Ber+04; ASL11], defined as

$$\Psi(\mathbf{M}) = \frac{1}{2}\sum_{i=1}^{R}\left(\sum_{\substack{j=1\\j\neq i}}^{R}\|\mathbf{m}_i - \mathbf{m}_j\|_2^2\right) = \frac{1}{2}\sum_{r=1}^{R}\|\mathbf{M}\mathbf{G}_r\|_{\mathrm{F}}^2 \tag{3.15}$$

where

$$\mathbf{G}_r = -\mathbf{I}_R + \mathbf{e}_r\mathbf{1}_R^{\mathrm{T}} \tag{3.16}$$

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

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

$$\Upsilon(\mathbf{dM}_t) = \frac{1}{2}\|\mathbf{dM}_t - \mathbf{dM}_{t-1}\|_{\mathrm{F}}^2. \tag{3.17}$$

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 3.4.3.

3.4. A two-stage stochastic program

3.4.1. General principle of a two-stage stochastic program

The following lines briefly recall the main ideas presented in the introduction of [RX11]. A two-stage stochastic program is generally expressed as

$$\min_{\mathbf{M}} \mathbb{E}_{\mathbf{Y},\mathbf{Z}}\big[f(\mathbf{Y},\mathbf{M},\mathbf{Z})\big] \text{ s.t. } \mathbf{M}\in\mathcal{M},\ \mathbf{Z}\in\mathcal{Z}. \tag{3.18}$$

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 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

$$\min_{\mathbf{Z}\in\mathcal{Z}} f(\mathbf{Y},\mathbf{M},\mathbf{Z}). \tag{3.19}$$

Given an independent and identically distributed (i.i.d) T-sample $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$, problem (3.18) can be approximated by the sample average approximation (SAA)

$$\min_{\mathbf{M}, \mathbf{Z}_1, \ldots, \mathbf{Z}_T} \frac{1}{T} \sum_{t=1}^{T} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{Z}_t), \text{ s.t. } \mathbf{M} \in \mathcal{M},\ \mathbf{Z}_t \in \mathcal{Z}. \tag{3.20}$$

Moreover, when the second-stage (3.19) admits a unique solution, (3.20) can be rewritten as

$$\min_{\mathbf{M} \in \mathcal{M}} \frac{1}{T} \sum_{t=1}^{T} h(\mathbf{Y}_t, \mathbf{M}) \tag{3.21}$$

$$h(\mathbf{Y}_t, \mathbf{M}) = \min_{\mathbf{Z} \in \mathcal{Z}} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{Z}) \tag{3.22}$$

which is the SAA corresponding to

$$\min_{\mathbf{M} \in \mathcal{M}} \mathbb{E}_{\mathbf{Y}} \big[h(\mathbf{Y}, \mathbf{M}) \big] \tag{3.23}$$

$$h(\mathbf{Y}, \mathbf{M}) = \min_{\mathbf{Z} \in \mathcal{Z}} f\big(\mathbf{Y}, \mathbf{M}, \mathbf{Z}\big) \tag{3.24}$$

where the two stages explicitly appear. However, f defined in (3.7) is non-convex with respect to $\mathbf{Z} = (\mathbf{A}, \mathbf{dM})$, where $\mathcal{Z} = \mathcal{A}_R \times \mathcal{D}$. Thus, problem (3.19) does not necessarily 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 [RX11]. This framework enables a convergence result in terms of a critical point $\{\mathbf{M}, \mathbf{Z}_1, \ldots, \mathbf{Z}_T\}$ of (3.20) to be obtained. However, the significant size of the SAA problem (3.20) in our case is generally too expensive from a computational point of view. To address this issue, we propose to slightly adapt the work developed in [Mai+10] to propose an online estimation algorithm described in Algo. 5. This algorithm has the same convergence property as [Mai+10] 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 3.4.3.

Algorithm 5: 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 [BST13], cf. §3.4.2
b           $(\mathbf{A}_t, \mathbf{dM}_t) \in \arg\min_{(\mathbf{A}, \mathbf{dM}) \in \mathcal{A}_R \times \mathcal{D}_t} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}, \mathbf{dM})$;
            $\mathbf{C}_t \leftarrow \xi \mathbf{C}_{t-1} + \mathbf{A}_t \mathbf{A}_t^T$;
            $\mathbf{D}_t \leftarrow \xi \mathbf{D}_{t-1} + (\mathbf{dM}_t \mathbf{A}_t - \mathbf{Y}_t) \mathbf{A}_t^T$;
            $\mathbf{E}_t \leftarrow \xi \mathbf{E}_{t-1} + \mathbf{dM}_t$;
            // Endmember update [Mai+10, Algo. 2], cf. §3.4.2
c           $\mathbf{M}^{(t)} \leftarrow \arg\min_{\mathbf{M} \in \mathcal{M}} \hat{g}_t(\mathbf{M})$;
Result: $\mathbf{M}^{(T)}, \{(\mathbf{A}_t, \mathbf{dM}_t)\}_{t=1,\cdots,T}$
```

3.4.2. Parameter estimation

Whenever an image $\mathbf{Y}_t$ has been received, the abundances and variability are estimated by the PALM algorithm [BST13], which is guaranteed to converge to a critical point $(\mathbf{A}_t, \mathbf{dM}_t)$ of $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot) + \iota_{\mathcal{A}_\mathcal{R} \times \mathcal{D}_t}$. The endmembers are then updated by proximal gradient descent steps, similarly to [Mai+10]. Further details on the projections involved in this section are given in Appendix C.1.

Abundance and variability estimation. A direct application of [BST13] under the constraints (3.3) leads to the following abundance update rule

$$\mathbf{A}_t^{(k+1)} = \text{prox}_{\iota_{\mathcal{A}_R}} \left(\mathbf{A}_t^{(k)} - \frac{1}{L_{\mathbf{A}_t}^{(k)}} \nabla_{\mathbf{A}} f\big(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(k)}, \mathbf{dM}_t^{(k)}\big) \right) \tag{3.25}$$

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

$$\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \alpha(\mathbf{A}_t - \mathbf{A}_{t+1}) + (\mathbf{M}^{(t)} + \mathbf{dM}_t)^{\mathrm{T}} \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big], \text{ for } t = 1 \tag{3.26}$$

$$\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \alpha(2\mathbf{A}_t - \mathbf{A}_{t-1} - \mathbf{A}_{t+1}) + (\mathbf{M}^{(t)} + \mathbf{dM}_t)^{\mathrm{T}} \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big] \text{ for, } 1 < t < T \tag{3.27}$$

$$\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \alpha(\mathbf{A}_t - \mathbf{A}_{t-1}) + (\mathbf{M}^{(t)} + \mathbf{dM}_t)^{\mathrm{T}} \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big], \text{ for } t = T \tag{3.28}$$

$$L_{\mathbf{A}_t}^{(k)} = \left\| (\mathbf{M}^{(t)} + \mathbf{dM}_t^{(k)})^{\mathrm{T}} (\mathbf{M}^{(t)} + \mathbf{dM}_t^{(k)}) + \alpha\big(1 + [1 < t < T]\big)\mathbf{I}_R \right\|_{\mathrm{F}} \tag{3.29}$$

where $[\cdot]$ is the Iverson bracket introduced in the previous chapter. Note that the projection $\text{prox}_{\iota_{\mathcal{A}_R}}$ can be exactly computed using the algorithms proposed in [Duc+08; Con15]. Similarly, the update rule for the variability terms is

$$\mathbf{dM}_t^{(k+1)} = \text{prox}_{\iota_{\mathcal{D}_t}} \left(\mathbf{dM}_t^{(k)} - \frac{1}{L_{\mathbf{dM}_t}^{(k)}} \nabla_{\mathbf{dM}} f\big(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t^{(k+1)}, \mathbf{dM}_t^{(k)}\big) \right) \tag{3.30}$$

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

$$\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \gamma(\mathbf{dM}_t - \mathbf{dM}_{t+1}) + \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big] \mathbf{A}_t^{\mathrm{T}}, \text{ for } t = 1 \tag{3.31}$$

$$\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \gamma(2\mathbf{dM}_t - \mathbf{dM}_{t-1} - \mathbf{dM}_{t+1}) + \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big] \mathbf{A}_t^{\mathrm{T}}, \text{ for } 1 < t < T \tag{3.32}$$

$$\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \mathbf{M}^{(t)}, \mathbf{A}_t, \mathbf{dM}_t) = \gamma(\mathbf{dM}_t - \mathbf{dM}_{t-1}) + \big[(\mathbf{M}^{(t)} + \mathbf{dM}_t)\mathbf{A}_t - \mathbf{Y}_t \big] \mathbf{A}_t^{\mathrm{T}}, \text{ for } t = T \tag{3.33}$$

$$L_{\mathbf{dM}_t}^{(k)} = \left\| \mathbf{A}_t^{(k+1)} \mathbf{A}_t^{(k+1)\mathrm{T}} + \gamma\big(1 + [1 < t < T]\big)\mathbf{I}_R \right\|_{\mathrm{F}}. \tag{3.34}$$

Note that the projection $\text{prox}_{\iota_{\mathcal{D}_t}}$ can be efficiently approximated using the Dykstra algorithm (see [BD86; BCG11; HAS13]). The resulting algorithm is summarized in Algo. 6.

Algorithm 6: 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
        $\mathbf{A}_t^{(k+1)} \leftarrow \text{Update}\big(\mathbf{A}_t^{(k)}\big)$;    // cf. (3.25)
        // Variability update
        $\mathbf{dM}_t^{(k+1)} \leftarrow \text{Update}\big(\mathbf{dM}_t^{(k)}\big)$;    // cf. (3.30)
        $k \leftarrow k+1$;
    $\mathbf{A}_t \leftarrow \mathbf{A}_t^{(k)}, \mathbf{dM}_t \leftarrow \mathbf{dM}_t^{(k)}$;
Result: $(\mathbf{A}_t, \mathbf{dM}_t)$
```

Endmember estimation. Similarly, a direct application of the method detailed in [Mai+10; BST13] yields

$$\mathbf{M}^{(t,k+1)} = \text{prox}_{\iota_{\{\cdot \succeq \mathbf{0}\}}}\left(\mathbf{M}^{(t,k)} - \frac{1}{L_{\mathbf{M},t}} \nabla_{\mathbf{M}} \hat{g}_t\big(\mathbf{M}^{(t,k)}\big)\right) \tag{3.35}$$

where $L_{\mathbf{M},t}$ denotes the Lipschitz constant of $\nabla_{\mathbf{M}} \hat{g}_t(\mathbf{M}^{(t,k)})$ and

$$\nabla_{\mathbf{M}} \hat{g}_t(\mathbf{M}) = \mathbf{M}\left(\frac{1}{t}\mathbf{C}_t + \beta \sum_{r=1}^{R} \mathbf{G}_r \mathbf{G}_r^{\mathrm{T}}\right) - \frac{1}{t}\mathbf{D}_t \tag{3.36}$$

$$L_{\mathbf{M},t} = \left\| \frac{1}{t}\mathbf{C}_t + \beta \sum_r \mathbf{G}_r \mathbf{G}_r^{\mathrm{T}} \right\|_{\mathrm{F}}. \tag{3.37}$$

The resulting algorithm is summarized in Algo. 7.

Algorithm 7: 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
        $\mathbf{M}^{(t,k+1)} \leftarrow \text{Update}\big(\mathbf{M}^{(t,k)}\big)$;    // cf. (3.35)
        $k \leftarrow k+1$;
    $\mathbf{M}^{(t)} \leftarrow \mathbf{M}^{(t,k)}$;
Result: $\mathbf{M}^{(t)}$
```

3.4.3. Convergence guarantee

To ensure the convergence of the generated endmember sequence $\{\mathbf{M}^{(t)}\}_t$ towards a critical point of the problem (3.18), we make the following assumptions.

Assumption 1. The quadratic functions $\hat{g}_t$ are strictly convex, twice continuously differentiable, 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 $f(\mathbf{Y}_t, \mathbf{M}, \cdot, \cdot)$ 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 $\frac{\mu_\mathbf{M}}{2} \|\mathbf{M}\|_\text{F}^2$ 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.3, 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 3.3 are considered. Indeed, this assumption is obviously satisfied by the penalizations mentioned in this chapter. Assumption 3, which is crucial to Proposition 1, is further discussed in Appendix C.2 to ease the reading of this chapter.

Adapting the arguments used in [Mai+10], the convergence property summarized in Proposition 1 can be obtained.

Proposition 1 (Convergence of $\{\mathbf{M}^{(t)}\}_t$, [Mai+10])**.** *Under the assumptions 1,2 and 3, the distance between* $\mathbf{M}^{(t)}$ *and the set of critical points of the hyperspectral unmixing problem* (3.6) *converges almost surely to 0 when t tends to infinity.*

Proof. See sketch of proof reported in Appendix C.3. ☐

Remark 3.2. The proposed algorithm can be analyzed within the general framework introduced in [CPR16]. Indeed, the implementation of Algo. 5 can be seen as a particular case of [CPR16], since the inner iterations lead to an essentially cyclic update rule of the mixture parameters. Indeed, all the mixture parameters are updated at least once every $K = \max(2N_\text{iter}^P + 1, N_\text{iter} + 2)$ iterations, where N_iter^P and N_iter denote the number of iterations used by Algo. 6 and 7 respectively. This analysis, which requires weaker assumptions on the problem structure (satisfied for the unmixing problem under study), allows a deterministic convergence guarantee to be recovered (see [CPR16, Theorem 3.1]).

3.4.4. Computational complexity

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

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

where $N_\text{iter}^D, N_\text{iter}^P, N_\text{iter}$ denote the numbers of iterations for the Dykstra algorithm involved in the variability projection (3.30)), 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 (Algo. 6) and the endmember updates (Algo. 7) can be parallelized if needed due to the separability of the objective function f considered (separability with respect to the column of the abundance matrix, and with respect to the rows of the endmember and variability matrices).

3.5. Experiments with synthetic data

The performance of the proposed method has been evaluated on three image sequences composed of 10 images of size 98×102, with 173 bands. The images of the different sequences correspond to linear mixtures of $R \in \{3, 6, 10\}$

endmembers affected by smooth time-varying variability. The abundance maps associated with these simulation scenarii vary smoothly from one image to another, and do not necessarily satisfy the pure pixel assumption. The synthetic linear mixtures have been finally corrupted by an additive white Gaussian noise to ensure a resulting signal-to-noise ratio of SNR = $30\,\mathrm{dB}$.

As in chapter 2, the perturbed endmembers involved in the mixtures have been generated using the product of reference endmembers with randomly generated piecewise-affine functions. Some instances of the signatures used in the experiments are depicted in Fig. 3.1. Note that different affine functions have been considered at each time instant for each endmember.

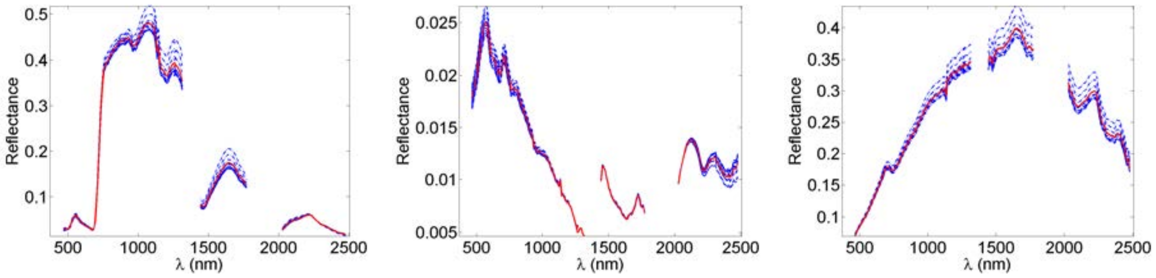


Figure 3.1.: Reference endmembers (red lines) and the corresponding instances under spectral variability (blue lines) involved in the synthetic HS images.

3.5.1. Compared methods

The results of the proposed algorithm have been compared to those obtained with classical linear unmixing methods individually applied to each image of the time series. The DSU [HCJ16] and PALM [BST13], which are batch algorithms, have been used as a reference whenever they were applicable. The methods compared in this chapter are recalled below with their most relevant implementation details. Note that the VCA/FCLS algorithm has been systematically used to initialize the compared methods when necessary.

1. VCA/FCLS (no variability): for each image, the endmembers are first extracted using the vertex component analysis (VCA) [NB05b] 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) by ADMM [BF10];
2. SISAL/FCLS (no variability): the endmembers are first extracted using the simplex identification via split augmented Lagrangian (SISAL) [Bio09]. Note that the pure pixel assumption is not required by 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 [Qia+11] 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 according to the recommendation given in [Qia+11];
4. PLMM: the BCD/ADMM algorithm described in chapter 1 is applied to each image with a stopping criterion set to 10^{-4}. The endmember regularization recalled in (3.15) has been used, with a parameter set to the same value as

Table 3.1.: Parameters used in the experiments.

(a) OU parameters.

	Synthetic data	Real data
ν^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

(b) ADMM parameters used for the DSU method [HCJ16] on the synthetic data.

	$R = 3$	$R = 6$	$R = 10$
τ^{incr}	1.1	1.1	1.1
τ^{decr}	1.1	1.1	1.1
μ	10	10	10
ρ	100	$10-1$	10^{-5}
ε^{abs}	10^{-2}	10^{-2}	10^{-2}
ε^{rel}	10^{-4}	10^{-4}	10^{-4}
N_{iter}	100	20	20

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 online unmixing (OU): 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. Algo. 6 (inner PALM algorithm) is stopped after N_{iter}^{P} iterations and the Dykstra algorithm used to compute the projection in (3.30) is iterated N_{iter}^{D} times. Moreover, Algo. 7 is stopped after N_{iter} iterations. Finally, Algo. 5 is stopped after N_{epochs} cycles – referred to as epochs – on the randomly permuted training set to approximately obtain i.i.d. samples [Mai+10]. In particular, the number of cycles N_{epochs} and sub-iterations N_{iter} have been empirically 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;

6. PALM [BST13]: a batch PALM algorithm has been applied to the synthetic datasets to obtain a reference in the ideal situation where all the data fit in memory. This algorithm is based on the model used in the previous chapter, in which the outlier term is omitted. The method has been initialized with the same initial points as the OU algorithm, for the same abundance, endmember and variability penalizations. The algorithm has been stopped when the relative difference of the objective function between two iterations is lower than 10^{-5}, with a maximum of 500 iterations;

7. DSU [HCJ16]: the DSU algorithm (considered in chapter 2) has been applied to the entire dataset with the same initialization as the batch PALM. The parameters used for this method are given in Table 3.2b. To ensure a fair comparison with the other methods, the variability extracted at time instant t is interpreted as the deviation of the endmember matrix extracted for the image t from the endmember temporal average.

The performance of the compared methods, which has been assessed in terms of the criteria introduced in the preliminary chapter (see section I.4), is reported in the following section.

3.5.2. Results

The values for the parameters of the proposed approach have been chosen using a grid of values successively tested for each parameter while the others are fixed (as in chapter 1). The value selected for each parameter corresponds to the one minimizing the reconstruction error RE. The values taken to initialize this procedure ensure that each penalization term has a weight corresponding to a predefined fraction of the initial data fitting term (typically 10 to 20 %). A more detailed study of the influence of these parameters on the reconstruction error can be found in Appendix C.4.

The values selected for the parameters involved in the proposed algorithm are given in Table 3.1. For the dataset associated with mixtures of $R = 3$ endmembers, the abundance maps obtained by the proposed method are compared to those of the other unmixing methods in Figs. 3.2 to 3.4, whereas the corresponding endmembers are displayed in Fig. 3.5. The performance of the different algorithms, reported in Table 3.3, leads 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 PLMM 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 (last line of Figs. 3.2 to 3.4) are closer to the ground truth (first line) than VCA/FCLS (third line). This observation is confirmed by the results given in Table 3.3;

- The proposed method provides competitive unmixing results while allowing temporal endmember variability to be estimated for each endmember (see Fig. 3.5). Since the problem considered is not convex, we can observe that the results obtained by the batch PALM algorithm and the DSU method are not systematically better than those of the online algorithm;

- 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 3.3), without introducing much more degrees of freedom into the underlying model when compared to PLMM;

- Even though the performance of the proposed method degrades with the number of endmembers, the results remain better than, 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. 3.2 to 3.4. 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. 3.2 to 3.4, to assess the difficulty of recovering each endmember in each image. This result was expected, since VCA is a pure pixel-based unmixing method.

Table 3.3.: Simulation results on synthetic data (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$).

		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	aSAM($\mathbf{Y}$) (°)	time (s)
$R = 3$	VCA	16.8	4.20	/	0.37	2.81	1.4
	SISAL	16.5	3.83	/	0.35	2.75	3
	$\ell_{1/2}$ NMF	19.4	7.39	/	0.77	3.1	189
	PLMM	17.2	4.22	**0.65**	**0.12**	**1.53**	380
	OU	4.70	0.27	2.07	0.34	2.75	156
	PALM	5.02	**9.67**$\times 10^{-3}$	1.81	0.34	2.75	37
	DSU [HCJ16]	**2.87**	0.35	1.74	3.57	2.76	24
$R = 6$	VCA	2.50	0.08	/	1.89	2.09	5
	SISAL	2.46	0.76	/	1.32	1.82	6
	$\ell_{1/2}$ NMF	**2.15**	0.07	/	1.38	1.85	402
	PLMM	2.43	0.10	2.85	**0.57**	**1.19**	783
	OU	2.44	0.22	8.67	1.33	1.83	315
	PALM	2.47	**6.21** $\times 10^{-3}$	10.04	1.33	1.82	367
	DSU [HCJ16]	2.25	0.43	4.09	11.84	1.82	10
$R = 10$	VCA	3.82	2.72	/	386.0	4.6	5
	SISAL	10.0	3.43	/	462.0	16.3	8
	$\ell_{1/2}$ NMF	5.36	2.88	/	422.0	5.95	254
	PLMM	5.72	1.92	**6.51**	382.0	4.68	3567
	OU	3.32	0.47	13.00	1.98	2.11	191
	PALM	**2.89**	**7.19** $\times 10^{-3}$	13.59	**1.93**	2.10	218
	DSU [HCJ16]	3.99	1.47	7.11	14.75	**1.97**	52

Figure 3.2.: Abundance maps of the first endmember used in the synthetic mixtures. 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.

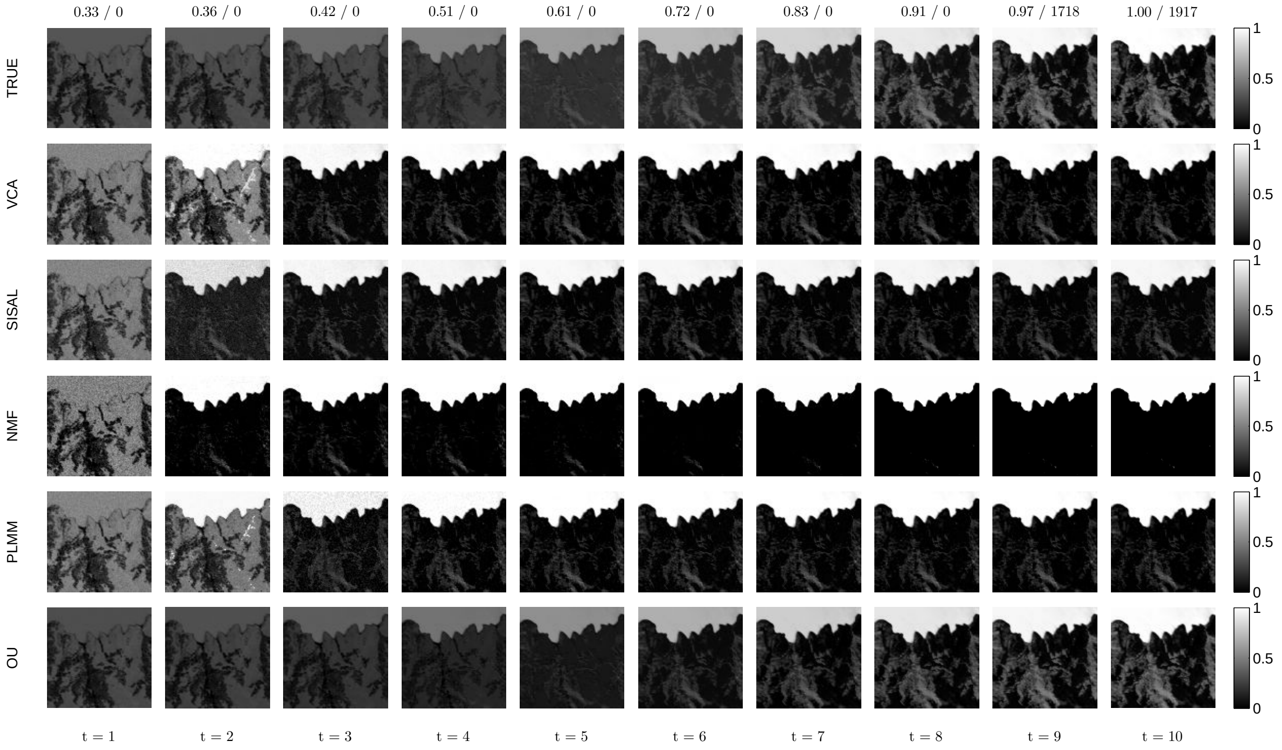
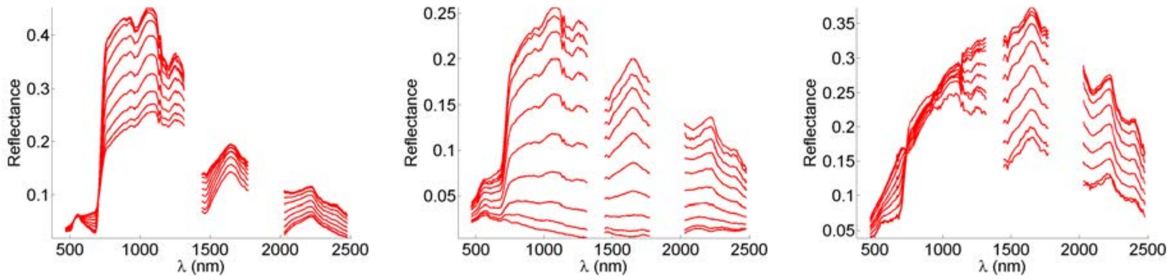
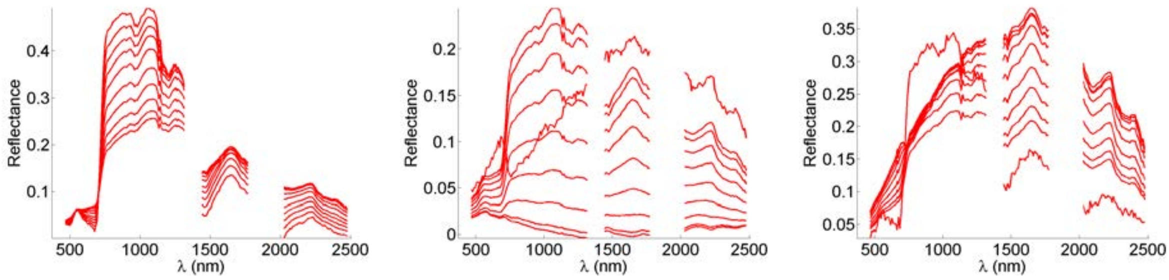


Figure 3.3.: Abundance maps of the second endmember used in the synthetic mixtures. The top line indicates the theoretical maximum abundance value and the true number of pixels whose abundance is greater than 0.95 for each time instant.

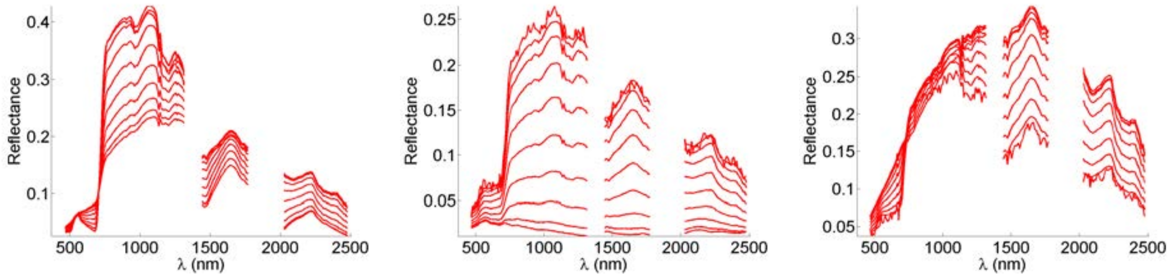
Figure 3.4.: Abundance maps of the third endmember used in the synthetic mixtures. The top line indicates the theoretical maximum abundance value and the true number of pixels whose abundance is greater than 0.95 for each time instant.



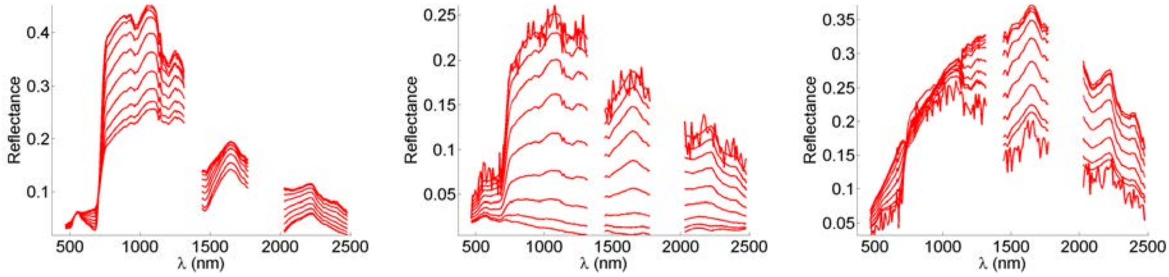
(a) VCA (b) VCA (c) VCA



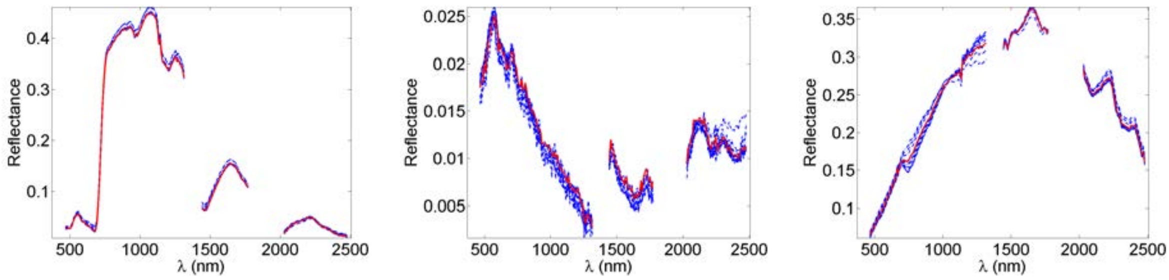
(d) SISAL (e) SISAL (f) SISAL



(g) NMF (h) NMF (i) NMF



(j) PLMM (k) PLMM (l) PLMM



(m) OU (n) OU (o) OU

Figure 3.5.: Estimated endmembers on the synthetic hyperspectral time series (extracted endmembers are represented in red, variability in blue dotted lines).

3.6. Experiments with real data

3.6.1. Description of the dataset

The proposed algorithm has been applied to real AVIRIS HS images acquired over the Mud Lake located in the Lake Tahoe region (California, United States of America) between 2014 and 2015[1]. 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 [BN08], those of the recently proposed eigen-gap approach (EGA) [Hal+16b] (see Table 3.4), and the consistency of the resulting abundance maps. The parameters used for the proposed approach are given in Table 3.1, and the other methods have been run with the same parameters as in section 3.6. 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 (see Fig. 3.6(e)). For this dataset, no comparison can be made with the batch PALM and the DSU algorithms, since the whole dataset does not fit in the memory of the computer used for the following experiments.

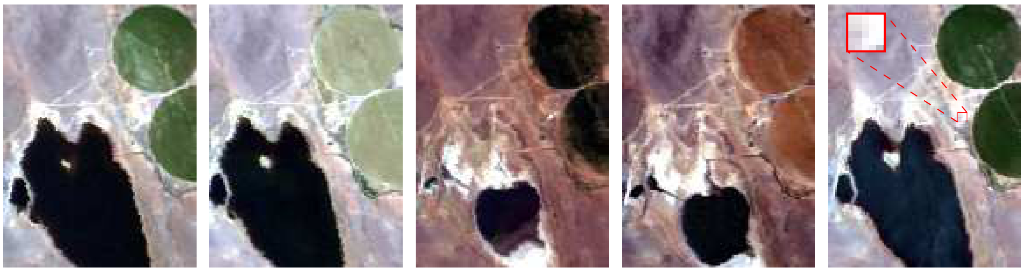

(a) 04/10/2014 (b) 06/02/2014 (c) 09/19/2014 (d) 11/17/2014 (e) 04/29/2015

Figure 3.6.: Scenes used in the experiment, given with their respective acquisition date. The area delineated in red in Fig. 3.6(e) highlights a region known to contain outliers, which has been removed prior to the unmixing procedure (this observation results from the analysis conducted in chapter 2).

Table 3.4.: Endmember number R estimated on each image of the real dataset by HySime [BN08] and EGA [Hal+16b].

	04/10/2014	06/02/2014	09/19/2014	11/17/2014	04/29/2015
HySime [BN08]	16	21	19	21	22
EGA [Hal+16b]	3	5	4	3	3

3.6.2. Results

Since no ground truth is available, the algorithm performance is evaluated in terms of the reconstruction error defined in section I.4. Only the more consistent abundance maps and endmembers obtained for $R = 3$ are presented in Figs. 3.7 to 3.10. The proposed method provides comparable reconstruction errors (see Table 3.5), 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

---

[1] The scene under study is located in the same area as the real dataset used in the experiments reported in chapter 2.

Fig. 3.9 (area delineated in red) presents significant errors when visually compared to the corresponding RGB image in Fig. 3.6(d). 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. 3.10(a) and 3.10(c) (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 the $\ell_{1/2}$ NMF and PLMM 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. 3.6(d)). Besides, PLMM yields very low reconstruction errors at the price of a computational cost which may become prohibitive for extended image sequences.

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. 3.6) – are found by the proposed method to be affected by the most significant variability level (see Table 3.6). In this experiment, a significant increase can be observed in the endmember variability energy over the last three images of the sequence (see Table 3.6), suggesting that the endmembers are apparently better represented in the two first images of the sequence (see Fig. 3.6). 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 study.

Table 3.5.: 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
	PLMM	**0.37**	2449
	Proposed (OU)	1.04	134
$R=4$	VCA/FCLS	43.8	**2**
	SISAL/FCLS	0.35	3
	$\ell_{1/2}$ NMF	16.0	163
	PLMM	**0.27**	4396
	Proposed (OU)	0.76	126
$R=5$	VCA/FCLS	63.9	**2**
	SISAL/FCLS	0.17	4
	$\ell_{1/2}$ NMF	14.6	174
	PLMM	**0.098**	12511
	Proposed (OU)	0.17	128

Table 3.6.: Experiment with real data for $R = 3$: energy of the variability captured for each endmember at each time instant ($\|\mathbf{dm}_{r,t}\|_2 / L \times 10^{-5}$ for $r = 1, \ldots, R$, $t = 1, \ldots, T$).

	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

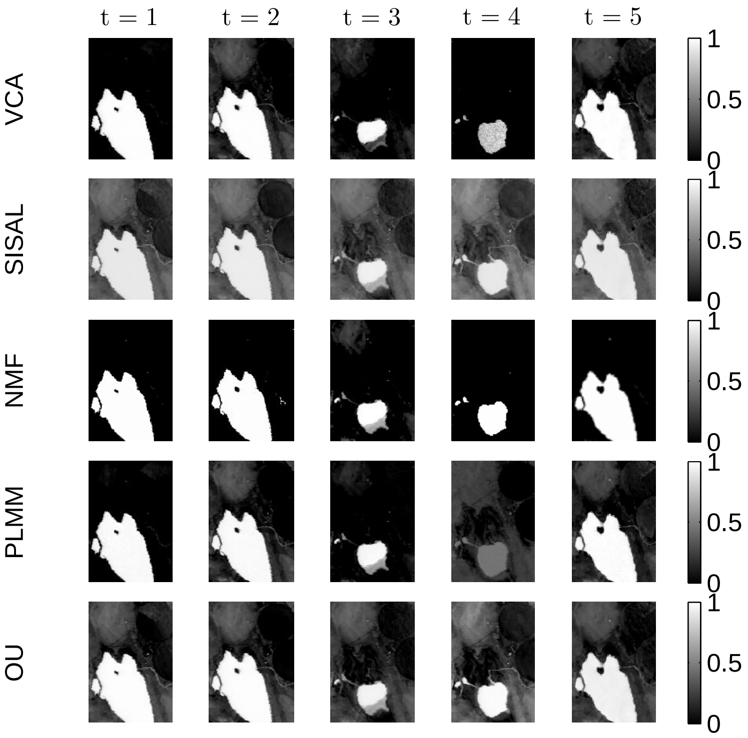


Figure 3.7.: Water abundance maps.

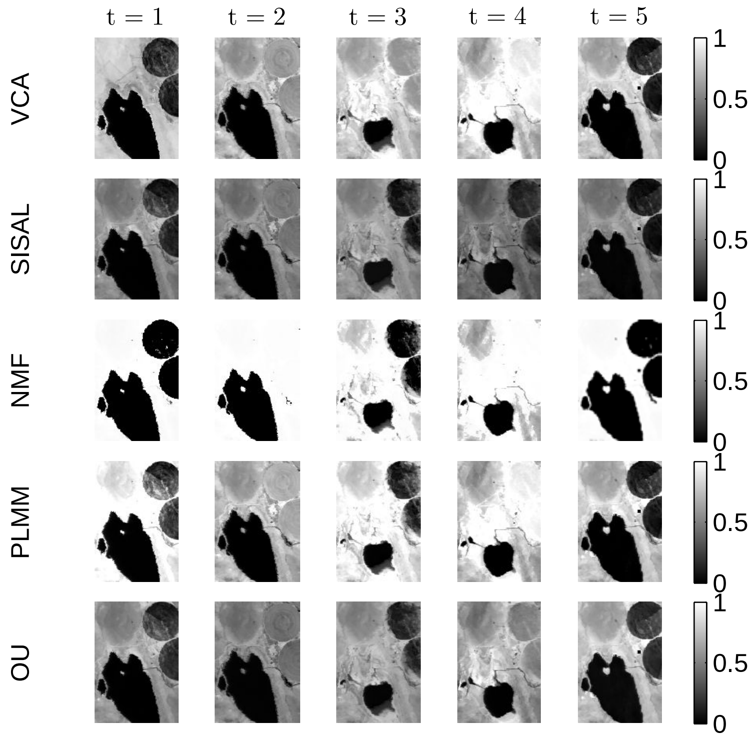


Figure 3.8.: Soil abundance maps.

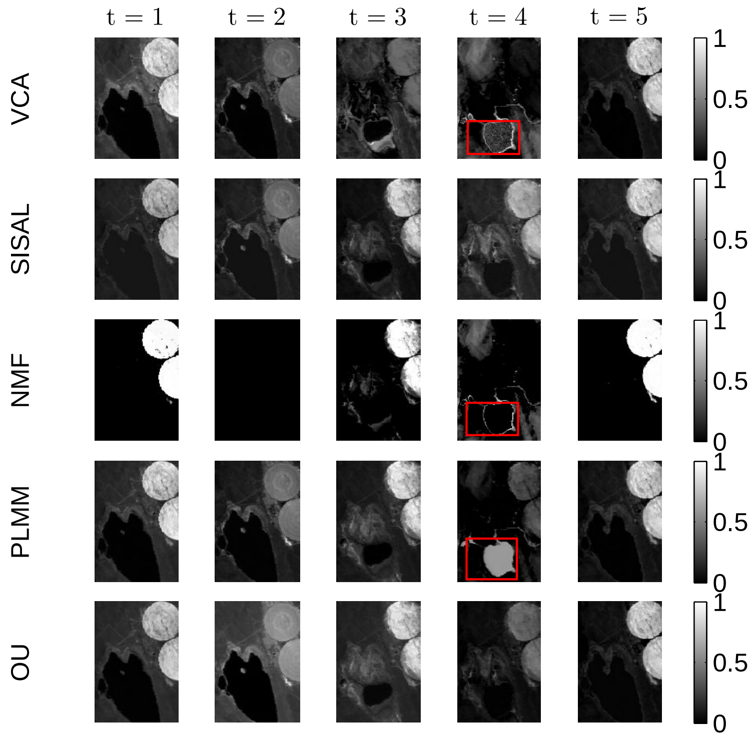
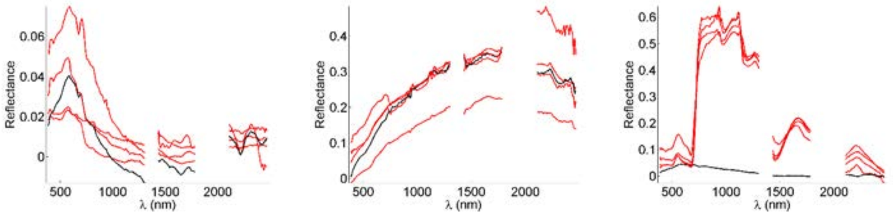
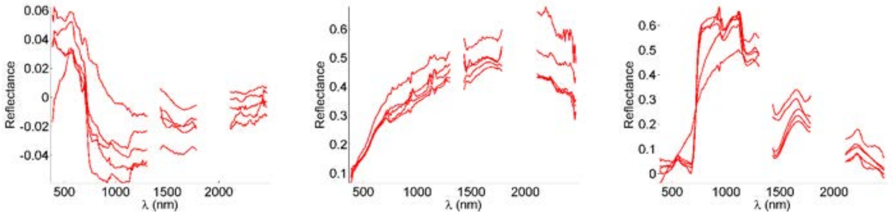


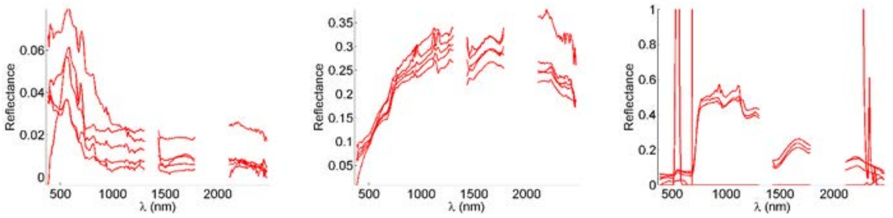
Figure 3.9.: Vegetation abundance maps. The region delineated in red, where almost no vegetation is supposed to be present, reveals that the water endmember extracted by VCA has been split into two parts. This endmember identification error has an influence on $\ell_{1/2}$ NMF and PLMM, but the error is relatively well corrected by the abundance sparsity prior used in $\ell_{1/2}$ NMF, at the price of signficantly degraded endmember estimation (see Fig. 3.10(i)). This observation is further confirmed in Figs. 3.10(a) and 3.10(c). The results obtained by $\ell_{1/2}$ NMF and PLMM are similarly affected, since initialized with VCA.



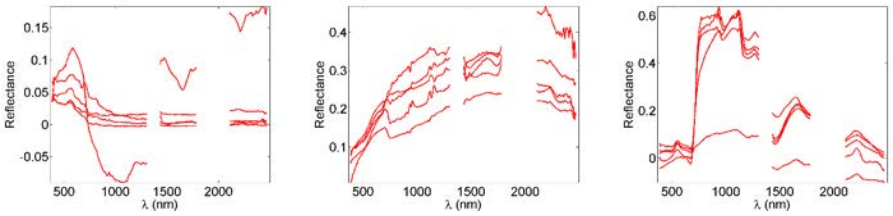
(a) Water (VCA) (b) Soil (VCA) (c) Vegetation (VCA)



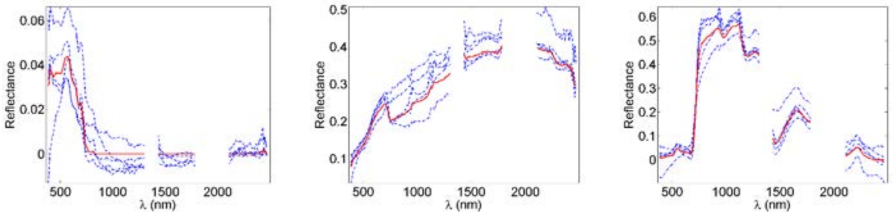
(d) Water (SISAL) (e) Soil (SISAL) (f) Vegetation (SISAL)



(g) Water (NMF) (h) Soil (NMF) (i) Vegetation (NMF)



(j) Water (PLMM) (k) Soil (PLMM) (l) Vegetation (PLMM)



(m) Water (OU) (n) Soil (OU) (o) Vegetation (OU)

Figure 3.10.: Endmembers ($\mathbf{m}_r$, red lines) and their variants affected by variability ($\mathbf{m}_r + \mathbf{dm}_{r,t}$, blue dotted lines) recovered by the different methods from the real dataset depicted in Fig. 3.6. Signatures corresponding to different time instants are represented in a single figure to better appreciate the variability recovered from the data. The spectra represented in black correspond to signatures identified by VCA on the image 3.6(d), where the water endmember has been split into two parts (see Figs. 3.10(a) and 3.10(c)). Note that the estimation of the vegetation signature by $\ell_{1/2}$ NMF is relatively poor (signature of amplitude significantly larger than 1) for the image 3.6(d), as illustrated in Fig. 3.10(i).

3.7. Conclusion and future work

This chapter introduced an online unmixing procedure based on the perturbed linear model developed in chapter 1. The underlying unmixing problem, formulated as a two-stage stochastic program, was solved by an online algorithm. The proposed algorithm was designed to unmix multiple HS images of moderate dimensions, potentially affected by smoothly varying endmember perturbations. Indeed, the adopted cost function presents a number of spurious local optima which can significantly increase with the size of the images and the number of endmembers considered. This problem is however common to many blind-source separation problems.

Simulations conducted on synthetic and real data enabled the interest of the proposed 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.

In this respect, the last chapter of this manuscript investigates the implementation of a partially asynchronous unmixing algorithm, thus offering a complementary point of view to the question addressed in this chapter.

Main contributions. An online unmixing algorithm has been proposed to analyze multi-temporal hyperspectral images, whose size may preclude the use of batch unmixing procedures. Based on a modified version of the model introduced in chapter 1, the problem is formulated as a two-stage stochastic program, which is solved using a stochastic approximation. The results obtained on real data show the interest of considering several HS images to improve the unmixing results.

3.8. Conclusion (in French)

Ce chapitre a permis d’introduire un algorithme de démélange en ligne, basé sur une variante du modèle introduit au chapitre 1 et formulé sous la forme d’un problème d’optimisation stochastique avec recours. L’algorithme proposé est destiné à l’analyse de plusieurs images hyperspectrales de dimension modérée, affectées par une variabilité temporelle qui évolue lentement au cours du temps. En effet, la fonction coût adoptée en pratique comporte un nombre d’extrema locaux qui peut augmenter de façon significative avec la dimension du problème considéré. Ce problème est néanmoins commun à un grand nombre de méthodes de séparation de sources de la littérature. Les simulations numériques, conduites sur données synthétiques et réelles, ont permis d’évaluer la pertinence de l’approche proposée, en comparaison avec une analyse individuelle de chacune des images à disposition.

Par la suite, il conviendrait d’évaluer la robustesse de la technique proposée à l’égard d’une erreur d’estimation du nombre de source R, supposé fixé *a priori*. Une application de cette approche à des données réelles composées d’un plus grand nombre de sources, ou la prise en compte simultanée de la variabilité spatiale et temporelle pour des problèmes de détection de changements, constituent d'autres perspectives de recherche à court terme.

À ce titre, le développement d'un algorithme de démélange asynchrone est envisagé dans le chapitre suivant, offrant ainsi un point de vue complémentaire à celui développé dans ce chapitre.

Contributions principales. Une stratégie de démélange en ligne est proposée pour analyser une collection d'images hyperspectrales, acquises au-dessus d'une même scène d'intérêt à différents instants. Basé sur une variante du modèle introduit dans le permier chapitre, le problème de démélange est formulé sous la forme d'un problème d'optimisation stochastique avec recours. Les résultats obtenus sur données synthétiques et réelles montrent l'intérêt d'utiliser plusieurs images hyperspectrales pour améliorer les résultats du démélange.

CHAPTER 4

A partially asynchronous distributed unmixing algorithm

This chapter has been adapted from the submitted journal paper [TDT17b] .

Contents

4.1. Introduction (in French)

Les chapitres 2 et 3 ont permis d'illustrer l'intérêt de considérer plusieurs images hyperspectrales pour analyser une même scène d'intérêt. À ce titre, un algorithme de démélange en ligne a été proposé au chapitre 3 pour estimer les paramètres d'un modèle de mélange en présence de variabilité temporelle. En complément de cette approche, ce chapitre s'intéresse à la mise en œuvre d'une procédure distribuée pour traiter des images (ou une séquence d'images) de grande dimension.

Les algorithmes distribués, synchrones ou (partiellement) asynchrones, s'avèrent particulièrement adaptés à la résolution de problèmes de grande dimension [BJ13; Lia+14; CTS15; LS15; FSS15; Scu+17; Yan+16; PR15; CE16]. Pour un grand nombre d'applications, les tâches sont réparties par une unité de calcul maître (ou processus maître) entre plusieurs noeuds de calculs qui lui sont subordonnés. Dans le cas synchrone (Fig. 4.1(a)), le maître attend l'ensemble des informations traitées par les noeuds de calcul avant de les aggréger et de procéder à une nouvelle répartition des tâches. Dans le cas asynchrone, certains problèmes permettent de se dispenser d'une hiérachie entre les différentes unités disponibles : chacun des noeuds effectue une tâche donnée en synchronisant régulièrement ses résultats avec ceux de ses voisins dans le cadre d'un système de voisinage prédéfini [BHI16; Dav16; Can+16]. Pour d'autres applications (telles que celle traitée dans ce chapitre), le cas asynchrone permet de s'affranchir en partie de la contrainte de synchronisation en exploitant au mieux la vitesse de calcul relative de chacun des noeuds, ce qui permet de réduire leurs périodes d'inactivité [Pen+16; Dav16; Can+16]. De fait, l'unité maître assigne une nouvelle tâche à un noeud dès qu'il a terminé ses calculs (voir Fig. 4.1(b)). Dans ce contexte partiellement asynchrone, l'information disponible en chacun des noeuds ne présente pas forcément un même niveau de traitement [CE16].

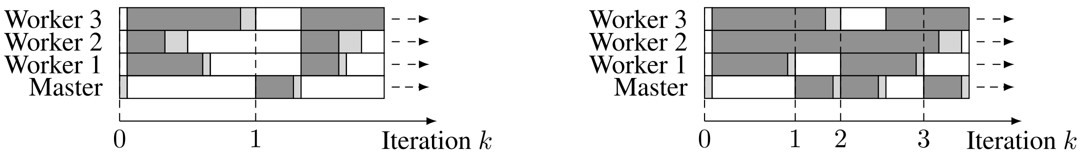
Ce dernier chapitre est organisé comme suit. Le problème d'optimisation traité dans ce chapitre ainsi que la procédure distribuée retenue pour le problème de démélange sont introduits de façon générale dans la section 4.3, avec l'ensemble des hypothèses qui permettront de bénéficier de résultats de convergence établis dans la littérature. Une application au problème de démélange linéaire, en présence ou non de variabilité, est proposée dans le section 4.4. Les résultats de simulation sur données synthétiques et réelles, rapportés aux sections 4.5 et 4.6, permettent d'évaluer l'apport de l'asynchronicité par rapport à une version synchrone de l'algorithme considéré. La section 4.7 permet de conclure cette étude et d'y apporter plusieurs perspectives de recherche.

4.2. Introduction

The preceding chapters have shown the interest of unmixing multiple HS images to better characterize the scene under study, and an online unmixing algorithm has been proposed to address the resulting optimization problem. In this chapter, a complementary approach to the study presented in chapter 3 is considered. More precisely, a distributed procedure is investigated to estimate the mixture parameters when considering large – or sequences of – HS images.

In practice, distributed algorithms are of a particular interest to address large scale optimization problems [BJ13; Lia+14; CTS15; LS15; FSS15; Scu+17; Yan+16; PR15; CE16]. In the case of synchronous distributed procedures (see Fig. 4.1(a)), a master node waits for the completion of operations conducted by several independent computation nodes (workers) before proceeding to the next iteration. More precisely, each iteration of the master consists in aggregating information received from the workers, and then assigning a new task to each worker.

Asynchronous algorithms offer more flexibility in terms of synchronization requirement. Indeed, they exploit the difference in the computation time required by each worker, thus reducing idleness periods [Pen+16; Dav16; Can+16]. For some applications, there is no hierarchy between the computation nodes, and each worker can be triggered at any time, independently from the other nodes [BHI16; Dav16; Can+16]. In the partially asynchronous setting considered in this chapter, a master node first assigns different tasks to all the available workers, then aggregates information from a given node as soon as it receives its information. The master finally relaunches a task on the node it has received information



(a) Synchronous distributed system

(b) Asynchronous distributed system

Figure 4.1.: Illustration of a synchronous and an asynchronous distributed mechanism (idle time in white, transmission delay in light gray, computation delay in gray). In the synchronous configuration, the master is triggered once it has received information from all the workers. In the asynchronous case, the master node is triggered whenever it has received information from K workers ($K = 1$ in the illustration).

from (see Fig. 4.1(b)). In this setting, the workers may make use of out-of-date information to perform their local update [CE16].

The literature devoted to distributed optimization methods includes a wide variety of approaches for both convex and non-nonconvex problems. Consequently, an exhaustive list of the existing methods is out of the scope of the present study. The emphasis will be specifically put on recently proposed asynchronous algorithms which can be of interest to matrix factorization problems, and more precisely to hyperspectral unmixing. In this perspective, we focus more specifically on non-convex problems of the form

$$\begin{aligned}
&(\mathbf{x}^*, \mathbf{z}^*) \in \underset{\mathbf{x},\mathbf{z}}{\arg\min}\, \Psi(\mathbf{x}, \mathbf{z}) \\
&\Psi(\mathbf{x}, \mathbf{z}) \triangleq \sum_{t=1}^{T} \Big[f_t(\mathbf{x}_t, \mathbf{z}) + \sum_{j=1}^{J} g_{jt}(\mathbf{x}_{jt}) \Big] + \sum_{i=1}^{I} r_i(\mathbf{z}_i), \quad \mathbf{x}_t = \big[\mathbf{x}_{j,t}\big]_j
\end{aligned} \tag{4.1}$$

whose resolution is conducted by (partially) asynchronous optimization algorithms.

The problem structure defined in (4.1) encompasses many problems, which can be efficiently distributed in terms of the *local* cost functions f_t. Each of these functions depends on a local variable $\mathbf{x}_t$, and a global variable $\mathbf{z}$ shared between the different functions f_t. Depending on the expression of f_t, the problem might be further decomposed into several blocks (i.e., $\mathbf{x}_{j,t}$ and $\mathbf{z}_i$). Hereafter, f_t represents a smooth (non-necessarily convex) data fitting term, whereas g_{jt} and r_i are penalizations whose expression is problem-dependent. To address matrix factorization problems, we propose to consider a centralized computing architecture as in [Cha+16], in which a master unit is in charge of the shared variable $\mathbf{z}$, and supervises T workers which have access to a variable $\mathbf{x}_t$ and a (possibly out-of-date) copy of $\mathbf{z}$. This setting can be notably seen as a particular instance of the general framework described in [Can+16].

Several asynchronous methods recently proposed in the literature are adapted to the problem of interest [Lia+15; Cha+16; Pen+16; Dav16; Can+16], at least in theory. In particular, Gauss-Seidel optimization schemes appear as a convenient technique to decompose the original problem into a collection of simpler optimization tasks, which can be conducted and distributed efficiently [Boy+10; WYZ15]. We can mention the partially asynchronous ADMM proposed in [Cha+16], used to solve a similar problem reformulated as a consensus optimization task. However, hyperspectral unmixing in particular does not lend itself well to such approaches, since most of the involved subproblems require the use of iterative solvers. In this case, the proximal alternating linearized minimization (PALM) [BST13] and its extensions

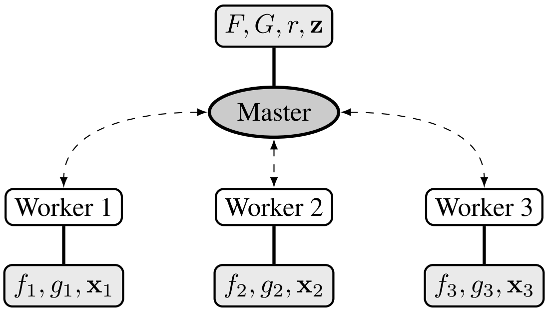


Figure 4.2.: Illustration of the master-slave architecture considered for the problem (4.2) with $T = 3$ workers (the function and variables available at each node are given in light gray rectangles).

[FGP15; CPR16] offer the advantage to benefit from appropriate convergence guarantees in non-convex settings while allowing parts of the computations to be easily distributed. In addition, recent works have investigated the development of an asynchronous version of the PALM algorithm [Dav16]. However, the descent steps prescribed in [Dav16] to ensure the algorithm convergence explicitly depend on a maximum allowable delay τ, which is generally unknown and system-dependent. Besides, the residual terms resulting from the allowed asynchronicity lead to extremely small admissible step-sizes, and do not result in any practical improvement in terms of convergence speed (see for instance [Can+16] and the numerical experiments reported in section 4.5). From this point of view, [Cha+16] appears more convenient in the sense that the maximum allowable delay is explicitly controlled. However, resorting to an ADMM algorithm suffers from the previously mentioned drawback, and does not ensure that the constraints imposed on the variables are satisfied at each iteration. Finally, the strategy in [Can+16] allows more flexibility in the step-size when compared to [Dav16].

Given the preceding remarks, the framework introduced in [Can+16] appears as one of the most flexible to address matrix factorization problems. We consequently propose to resort to the framework introduced in [Can+16], which encompasses the structure described in [Cha+16], to address hyperspectral unmixing, i.e., a particular matrix factorization problem. This choice is partly justified by the possible connections between the PALM algorithm and [Can+16]. Indeed, the PALM algorithm enables a synchronous distributed algorithm to be easily derived for the problem studied in chapter 1, which then offers an appropriate reference to precisely evaluate the relevance of the asynchronicity tolerated by [Can+16].

The rest of this chapter is organized as follows. The assumptions on the problem structure are introduced in section 4.3. The expression of this problem essentially reflects the perspective of solving matrix factorization tasks. Numerical experiments conducted on hyperspectral unmixing problems is then proposed in section 4.4. Simulation results associated with this scenario illustrate the performance of the proposed approach on synthetic and real data in sections 4.5 and 4.6. Conclusions and future reasearch perspectives are outlined in section 4.7.

4.3. Proposed algorithm

Without loss of generality and to alleviate notations, we address the following optimization problem instead of (4.1)

$$(\mathbf{x}^*, \mathbf{z}^*) \in \underset{\mathbf{x},\mathbf{z}}{\arg\min}\, \Psi(\mathbf{x}, \mathbf{z}) \triangleq F(\mathbf{x}, \mathbf{z}) + G(\mathbf{x}) + r(\mathbf{z}) \tag{4.2}$$

with

$$F(\mathbf{x}, \mathbf{z}) = \sum_{t=1}^{T} f_t(\mathbf{x}_t, \mathbf{z}), \quad G(\mathbf{x}) = \sum_{t=1}^{T} g_t(\mathbf{x}_t). \tag{4.3}$$

In particular, the problem discussed in section 4.4 can be expressed under the form (4.2). With these notations, $\mathbf{x}_t$ denotes a *local* variable (i.e., which will be accessed in practice by a single worker), and $\mathbf{z}$ is a global variable (i.e., shared between the different workers, see Fig. 4.2). In fact, a convergence analysis for (4.1) can be easily derived from the analysis of (4.2) by induction on the number of blocks I and J. In the following, f_t denotes a (non-necessarily convex) smooth function in the sense precised in Assumption 5 (see section 4.3.2 below). Furthermore, g_t and r are assumed to be convex lower semi-continuous (non-necessarily smooth) functions whose proximal operator can be easily computed (i.e., no iterative solver is needed to approximate their computation). In practice, f_t plays the role of a data fitting term, whereas g_t and r can be regarded as regularizers or constraints. The structure of the algorithm investigated in this chapter, inspired by [Can+16], is briefly described in the following paragraph.

4.3.1. Algorithm description

Reminiscent of [Cha+16], the algorithm investigated in this chapter relies on a star topology configuration in which a master node supervises an optimization task distributed between several workers. The master node also updates and transmits a set of variables shared by the different workers. In fact, the computation time of synchronous algorithms is essentially conditioned by the speed of the slowest worker (see Figs. 4.1(a) and 4.1(b)). Consequently, relaxing the synchronization requirements (by allowing bounded delays between the information brought by each worker) allows a significant decrease in computation time to reach convergence, which can scale almost linearly with the number of workers [Cha+16; Dav16]. Note that, even though asynchronous optimization schemes may require more iterations than their synchronous counterparts to reach a given precision, allowing more frequent updates generally compensates this drawback in terms of computation time [Cha+16].

In the partially asynchronous setting, the master node updates the variables shared by the workers once it has received information from at least $K \ll T$ workers. The new state of the shared variable $\mathbf{z}$ is then transmitted to the K available workers, which can individually proceed to the next step. As in [Can+16], a relaxation step with decreasing step-sizes ensures the convergence of the algorithm (see Algo. 8). To alleviate notations and make clear to which extent the convergence analysis introduced in [Can+16] is applicable, we consider $K = 1$ in the rest of this chapter. Details on the operations performed by the master node and each worker are detailed in Algos. 8 and 9 respectively.

Note that a synchronous distributed counterpart of Algo. 8 can be easily derived for the problem (4.2), which partly justifies the form chosen for Algo. 8. This version consists in setting $\gamma_k = 1$, and waiting for the updates performed by all the workers ($K = T$, see Step 1 of Algo. 8) before updating the shared variable $\mathbf{z}$. This implementation will be taken as a reference to evaluate the computational efficiency of the proposed algorithm in sections 4.5 and 4.6.

```
Algorithm 8: Algorithm of the master node.
    Data: $\mathbf{x}^{(0)}$, $\mathbf{y}^{(0)}$, $\mathbf{z}^{(0)}$, $\gamma_0 \in (0,1]$, $\mu \in (0,1)$, $N_{\text{iter}}$, $K$.
    Broadcast $\mathbf{z}^{(0)}$ to the $T$ workers ;
    $k \leftarrow 0$ ;
    $\mathcal{T}_k \leftarrow \emptyset$ ;
    while $k < N_{iter}$ do
Step 1  Wait for $\hat{\mathbf{x}}_{t^k}^k$ from any worker ;
        $\mathcal{T}_k = \mathcal{T}_k \cup \{t^k\}$ ;
        $d_t^{k+1} = \begin{cases} 0 \text{ if } t \in \mathcal{T}_k \\ d_t^k + 1 \text{ otherwise} \end{cases}$ ;
        $\mathbf{x}_t^{k+1} = \begin{cases} \mathbf{x}_t^k + \gamma_k(\hat{\mathbf{x}}_t - \mathbf{x}_t^k) \text{ if } t \in \mathcal{T}_k \\ \mathbf{x}_t^k \text{ otherwise} \end{cases}$ ;
        if $(\sharp\mathcal{T}_k < K)$ then
            Go to step Step 1 ;                                  // wait until $\sharp\mathcal{T}_k \geq K$
        else
            $\hat{\mathbf{z}}^k \in \operatorname{prox}_{c_{\mathbf{z}}^k r}\left(\mathbf{z}^k + \frac{1}{c_{\mathbf{z}}^k}\nabla_{\mathbf{z}} F(\mathbf{x}^{k+1}, \mathbf{z}^k)\right)$;
            $\mathbf{z}^{k+1} = \hat{\mathbf{z}}^k + \gamma_k(\hat{\mathbf{z}}^k - \mathbf{z}^k)$;
            $\gamma_{k+1} = \gamma_k(1 - \mu\gamma_k)$;
            $\mathcal{T}_{k+1} \leftarrow \emptyset$ ;
            $k \leftarrow k + 1$;
    Result: $\mathbf{x}^{N_{\text{iter}}}$, $\mathbf{y}^{N_{\text{iter}}}$, $\mathbf{z}^{N_{\text{iter}}}$.
```

4.3.2. Convergence analysis

Based on the convergence results [BST13, Theorem 1] and [Can+16, Theorem 1], the proposed algorithm requires the following assumptions.

Assumption 4 (Algorithmic assumption). Let $(t_k, d_{t^k}^k) \in \{1, \ldots, T\} \times \{1, \ldots, \tau\}$ denote the couple composed of the index of the worker transmitting information to the master at iteration k, and the delay between the (local) copy $\tilde{\mathbf{z}}^k$ of the shared variable $\mathbf{z}$ and the current state $\mathbf{z}^k$ (i.e., $\tilde{\mathbf{z}}^k \triangleq \mathbf{z}^{k-d_{t^k}^k}$). The allowable delays $d_{t^k}^k$ are assumed to be bounded by a constant $\tau \in \mathbb{N}^*$. In addition, each couple $(t_k, d_{t^k}^k)$ represents a realization of a random vector within the probabilistic model introduced in [Can+16, Assumption C].

Assumption 5 (Inherited from PALM [BST13]). (i) For any $t \in \{1, \ldots, T\}$, $g_t : \mathbb{R}^n \to (-\infty, +\infty]$ and $r : \mathbb{R}^m \to (-\infty, +\infty]$ are proper, convex lower semi-continuous (l.s.c.) functions;

(ii) For $t \in \{1, \ldots, T\}$, $f_t : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a $\mathcal{C}^1$ function, and is convex with respect to each of its variables when the other is fixed;

(iii) Ψ, f_t, g_t, and r are lower bounded, i.e., $\inf_{\mathbb{R}^n \times \mathbb{R}^m} \Psi > -\infty$, $\inf_{\mathbb{R}^n \times \mathbb{R}^m} f_t > -\infty$, $\inf_{\mathbb{R}^n} g_t > -\infty$, and $\inf_{\mathbb{R}^m} r > -\infty$;

(iv) Ψ is a coercive semi-algebraic function (see [BST13]);

(v) For all $t \in \{1, \ldots, T\}$, $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{x}_t \mapsto f_t(\mathbf{x}_t, \mathbf{z})$ is a $\mathcal{C}^1$ function, and the partial gradient $\nabla_{\mathbf{x}_t} f_t(\cdot, \mathbf{z})$ is Lipschitz continuous with Lipschitz constant $L_{\mathbf{x}_t}(\mathbf{z})$. Similarly, $\mathbf{z} \mapsto f_t(\mathbf{x}_t, \mathbf{z})$ is a $\mathcal{C}^1$ function, and the partial gradient

Algorithm 9: Algorithm of the tth worker (since the shared variable $\mathbf{z}$ may have been updated by the master node in the meantime, $\tilde{\mathbf{z}}$ corresponds to a possibly delayed version of the current $\mathbf{z}^k$. From the master's point of view, $\tilde{\mathbf{z}} = \mathbf{z}^{k-d_t^k}$.

```
Data: $\tilde{\mathbf{z}}$, $\tilde{\mathbf{x}}_t$.
begin
    Wait for $(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_t)$ from the master node;
    $\hat{\mathbf{x}}_t \in \operatorname{prox}_{c_{\mathbf{x}_t} g_t}\left(\tilde{\mathbf{x}}_t - \frac{1}{c_{\mathbf{x}_t}} \nabla_{\mathbf{x}_t} f_t\left(\tilde{\mathbf{x}}_t, \tilde{\mathbf{z}}\right)\right)$;
    Transmit $\hat{\mathbf{x}}_t$ to the master node;
Result: $\hat{\mathbf{x}}_t$.
```

$\nabla_{\mathbf{z}} f_t(\mathbf{x}_t, \cdot)$ is Lipschitz continuous, with constant $L_{\mathbf{z},t}(\mathbf{x}_t)$;

(vi) the Lipschitz constants used in the algorithm, i.e., $L_{\mathbf{x}^k_{t^k}}(\tilde{\mathbf{z}}^k)$ and $L_{\mathbf{z},t_k}(\hat{\mathbf{x}}^k_{t^k})$ (denoted by $L^k_{\mathbf{x}^k_{t^k}}$ and $L^k_{\mathbf{z},t_k}$ in the following) are bounded, i.e. there exists appropriate constants such that for all iteration index k

$$0 < L^-_{\mathbf{x}} \leq L^k_{\mathbf{x}_{t^k}} \leq L^+_{\mathbf{x}}, \quad 0 < L^-_{\mathbf{z}} \leq L^k_{\mathbf{z},t^k} \leq L^+_{\mathbf{z}}.$$

(vii) ∇F is Lipschitz continuous on bounded subsets.

Assumption 6 (Additional assumptions). (i) For all $t \in \{1, \ldots, T\}$, $\mathbf{x}_t \in \mathbb{R}^n$, $\nabla_{\mathbf{x}_t} f_t(\mathbf{x}_t, \cdot)$ is Lipschitz continuous with Lipschitz constant $L_{\mathbf{x}_t,\mathbf{z}}(\mathbf{x}_t)$;

(ii) The Lipschitz constants $L_{\mathbf{x}_{t_k},\mathbf{z}}(\hat{\mathbf{x}}^k_{t^k})$ (denoted by $L^k_{\mathbf{x}_{t_k},\mathbf{z}}$ in the following) is bounded, i.e. there exists appropriate positive constants such that for all $k \in \mathbb{N}$:

$$0 < L^-_{\mathbf{x},\mathbf{z}} \leq L^k_{\mathbf{x}_{t_k},\mathbf{z}} \leq L^+_{\mathbf{x},\mathbf{z}}.$$

Assumption 4 summarizes standard algorithmic asumptions to ensure the convergence of Algo. 8. Besides, Assumption 5 gathers requirements of the traditional PALM algorithm [BST13], under which the distributed synchronous version of the proposed algorithm can be ensured to converge.

Under Assumptions 4 to 6, the analysis led in [Can+16] allows the following convergence result to be recovered.

Proposition 2 (adapted from [Can+16])**.** *Suppose that Problem* (4.2) *satisfies the requirements specified in Assumptions 4 to 6. Define the sequence* $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *of the iterates generated by Algos. 8 and 9, with* $\mathbf{v}^k \triangleq (\mathbf{x}^k, \mathbf{z}^k)$ *and the parameters in Algo. 9 chosen as*

$$c^k_{\mathbf{x}_{t^k}} = L^k_{\mathbf{x}_{t^k}}, \quad c^k_{\mathbf{z}} = L^k_{\mathbf{z}}.$$

Then, the following convergence results are obtained:

(i) the sequence $\{\Psi(\mathbf{v}^k)\}_{k\in\mathbb{N}}$ *converges almost surely;*

(ii) every limit point of the sequence $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *is a critical point of* Ψ *almost surely.*

Proof. See the sketch of proof proposed in Appendix D. □

The convergence analysis is conducted using an auxiliary function (introduced in Lemma 3 in Appendix) to handle asynchronicity [Dav16]. This analysis of this function finally allows convergence results associated with the original problem (4.2) to be recovered.

Besides, the following result ensures a stronger convergence guarantee for the synchronous counterpart of Algo. 8.

Proposition 3 (Finite length property, following from [BST13]). *Suppose that Problem* (4.2) *satisfies the requirements specified in Assumptions* 5 *to* 6*. Define the sequence* $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *of the iterates generated by the synchronous version of Algo.* 8*, with* $\mathbf{v}^k \triangleq (\mathbf{x}^k, \mathbf{z}^k)$ *and*

$$c^k_{\mathbf{x}_{t^k}} = L^k_{\mathbf{x}_{t^k}}, \quad c^k_{\mathbf{z}} = L^k_{\mathbf{z}}, \quad \gamma_k = 1, \quad K = T.$$

Then, the following properties can be proved:

(i) the sequence $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *has finite length;*

$$\sum_{k=1}^{+\infty} \|\mathbf{v}^{k+1} - \mathbf{v}^k\| < +\infty;$$

(ii) the sequence $\{\mathbf{v}^k\}_{k\in\mathbb{N}}$ *converges to a critical point of* Ψ*.*

Proof. These statements result from a direct application of [BST13, Theorem 1, Theorem 3] and [BST13, Remark 4 (iv)]. ☐

4.4. Application to hyperspectral unmixing

The approach described in the previous section is particularly adapted to address matrix factorization problems, and *a fortiori* hyperspectral unmixing. After briefly describing the application and notations used in the rest of this chapter, we introduce the optimization problem to be solved, and finally apply the algorithms described in the previous sections to estimate the parameters of interest.

4.4.1. Problem statement

Linear mixing model (LMM). For this study, we first consider the standard LMM introduced in paragraph I.2, where the data are divided into T blocks composed of N pixels[1]

$$\mathbf{Y}_t = \mathbf{M}\mathbf{A}_t + \mathbf{B}_t, \ t \in \{1, \dots, T\} \tag{4.4}$$

subject to the constraints

$$\mathbf{A}_t \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}_t^{\mathrm{T}} \mathbf{1}_R = \mathbf{1}_N, \quad \mathbf{M} \succeq \mathbf{0}_{L,R}. \tag{4.5}$$

---

[1] Note that each block can be composed of a different number of pixels if needed.

Assuming the data are corrupted by a white Gaussian noise leads to the following data fitting term

$$f_t(\mathbf{A}_t, \mathbf{M}) = \frac{1}{2} \left\| \mathbf{Y}_t - \mathbf{M}\mathbf{A}_t \right\|_{\mathrm{F}}^2 . \tag{4.6}$$

In addition, the constraints summarized in (4.5) are taken into account by defining

$$\begin{aligned}
g_t(\mathbf{A}_t) &= \iota_{\mathcal{A}_N}(\mathbf{A}_t)\\
\mathcal{A}_N &= \left\{ \mathbf{X} \in \mathbb{R}^{R \times N} \mid \mathbf{X}^{\mathrm{T}} \mathbf{1}_R = \mathbf{1}_N, \mathbf{X} \succeq \mathbf{0}_{R,N} \right\}\\
r(\mathbf{M}) &= \iota_{\{\cdot \succeq \mathbf{0}\}}(\mathbf{M})
\end{aligned}$$

leading to the optimization problem

$$(\mathbf{A}^*, \mathbf{M}^*) \in \underset{\mathbf{A}, \mathbf{M}}{\arg\min}\, \Psi(\mathbf{A}, \mathbf{M}). \tag{4.7}$$

The non-convex problem (4.7) satisfies Assumptions 5 to 6 (as a particular instance of the problem addressed in [BST13, Section 4]).

Perturbed LMM (PLMM). A similar application can be proposed with the variablity accounting model introduced in the first chapter. However, the non-negativity constraints imposed on the perturbed endmembers $\mathbf{M} + \mathbf{dM}_n$ should be removed in this context, since the workers do not have access to the latest version of the variable $\mathbf{M}$ when performing their local updates. The associated model is consequently defined as

$$\mathbf{Y}_t = \mathbf{M}\mathbf{A}_t + \underbrace{\left[\mathbf{dM}_{1,t}\mathbf{a}_{1,t}, \ldots, \mathbf{dM}_{N,t}\mathbf{a}_{N,t} \right]}_{\boldsymbol{\Delta}_t} + \mathbf{B}_t \tag{4.8}$$

subject to

$$\begin{aligned}
&\mathbf{A}_t \succeq \mathbf{0}_{R,N}, \quad \mathbf{A}_t^{\mathrm{T}} \mathbf{1}_R = \mathbf{1}_N\\
&\mathbf{M} \succeq \mathbf{0}_{L,R}, \quad \left\| \mathbf{dM}_{n,t} \right\|_{\mathrm{F}} \leq \nu, \ \forall (n,t) \in \{1, \ldots, N\} \times \{1, \ldots, T\}.
\end{aligned} \tag{4.9}$$

The optimization problem (4.2) can be easily adapted to this setting by defining

$$\begin{aligned}
\Psi(\underline{\mathbf{A}}, \mathbf{M}, \mathbf{d}\underline{\mathbf{M}}) &= F(\underline{\mathbf{A}}, \mathbf{M}, \mathbf{d}\underline{\mathbf{M}}) + r(\mathbf{M}) + \sum_t g_t(\mathbf{A}_t) + \sum_{n,t} h_{n,t}(\mathbf{dM}_{n,t})\\
f_t(\mathbf{A}_t, \mathbf{M}, \mathbf{dM}_t) &= \frac{1}{2} \left\| \mathbf{Y}_t - \mathbf{M}\mathbf{A}_t - \boldsymbol{\Delta}_t \right\|_{\mathrm{F}}^2\\
g_t(\mathbf{A}_t) &= \iota_{\mathcal{A}_N}(\mathbf{A}_t)\\
h_{n,t}(\mathbf{dM}_{n,t}) &= \iota_{\|\cdot\|_{\mathrm{F}} \leq \nu}(\mathbf{dM}_{n,t})
\end{aligned}$$

$$\mathcal{A}_N = \left\{ \mathbf{X} \in \mathbb{R}^{R\times N} \mid \mathbf{X}^{\mathrm{T}} \mathbf{1}_R = \mathbf{1}_N, \mathbf{X} \succeq \mathbf{0}_{R,N} \right\}$$
$$r(\mathbf{M}) = \iota_{\{\cdot \succeq \mathbf{0}\}}(\mathbf{M}).$$

Remark 4.1. In the initial formulation of the mixing model (4.4), the indexes t and T refer to subsets of pixels. A direct interpretation of this statement can be obtained by dividing a unique (and possibly large) hyperspectral image into T non-overlapping tiles of smaller (and possibly different) sizes. In this case, each tile is individually unmixed by a given worker. Another available interpretation allows multitemporal analysis to be conducted. Indeed, in practice, distributed unmixing procedures are of a particular interest when considering the unmixing of a sequence of several HS images, acquired by possibly different sensors at different dates, but sharing the same materials [HCJ16; YZP17]. In this case, t and T could refer to time instants. Each worker t is then dedicated to the unmixing of a unique HS image acquired at a particular time instant. The particular applicative challenge of distributed unmixing of multitemporal HS images partly motivates the numerical experiments on synthetic (yet realistic) and real data presented hereafter.

4.4.2. Parameter estimation

A direct application of [BST13] leads to the following update rule for the abundance matrix $\mathbf{A}_{t^k}$ of the model (4.4)

$$\widehat{\mathbf{A}}_{t^k}^k = \operatorname{prox}_{\iota_{\mathcal{A}_N}} \left(\mathbf{A}_t^k - \frac{1}{c_{\mathbf{A}_{t^k}}^k} \nabla_{\mathbf{A}_t} f_t \left(\mathbf{A}_{t^k}^k, \mathbf{M}^{k-d_{t^k}^k} \right) \right) \tag{4.10}$$

with

$$\nabla_{\mathbf{A}_t} f_t(\mathbf{A}_t, \mathbf{M}) = \mathbf{M}^{\mathrm{T}} \big(\mathbf{M}\mathbf{A}_t - \mathbf{Y}_t \big) \tag{4.11}$$

$$c_{\mathbf{A}_{t^k}}^k = L_{\mathbf{A}_{t^k}}^k = \left\| (\mathbf{M}^{k-d_{t^k}^k})^{\mathrm{T}} \mathbf{M}^{k-d_{t^k}^k} \right\|_{\mathrm{F}}. \tag{4.12}$$

Note that the projection $\operatorname{prox}_{\iota_{\mathcal{A}_N}}$ can be exactly computed (see [Duc+08; Con15] for instance). Similarly, the update rule for the endmember matrix $\mathbf{M}$ is

$$\widehat{\mathbf{M}}^k = \operatorname{prox}_{\iota_{\{\cdot \succeq \mathbf{0}\}}} \left(\mathbf{M}^k - \frac{1}{c_{\mathbf{M}}^k} \nabla_{\mathbf{M}} F \big(\mathbf{A}^{k+1}, \mathbf{M}^k \big) \right) \tag{4.13}$$

with

$$\nabla_{\mathbf{M}} F \big(\mathbf{A}, \mathbf{M} \big) = \sum_t (\mathbf{M}\mathbf{A}_t - \mathbf{Y}_t) \mathbf{A}_t^{\mathrm{T}} \tag{4.14}$$

$$c_{\mathbf{M}}^k = L_{\mathbf{M}}^k = \left\| \sum_t \mathbf{A}_t^{k+1} (\mathbf{A}_t^{k+1})^{\mathrm{T}} \right\|_{\mathrm{F}}. \tag{4.15}$$

Note that similar update rules can be derived for the model (4.9).

4.5. Experiments with synthetic data

To illustrate the interest of the allowed asynchronicity, we compare the estimation performance of Algo. 8 to the performance of its synchronous counterpart (described in section 4.3), and evaluate the resulting unmixing performance in comparison with three unmixing methods proposed in the literature. We propose to consider the context of multitemporal HS unmixing, which is of a particular interest for recent remote sensing applications [HCJ16; YZP17]. In this particular case, a natural way of distributing the data consists in assigning a single HS image to each worker. To this end, we generated synthetic data composed of $T = 3$ HS images resulting from linear mixtures of $R \in \{3, 6, 9\}$ endmembers acquired in $L = 413$ bands, with and without spectral variability (see chapter 1). The generated abundance maps vary smoothly over time (i.e., from one image to another) to reproduce a realistic evolution of the scene of interest. Each image, composed of $10,000$ pixels, was then corrupted by an additive white Gaussian noise whose variance ensures a signal-to-noise ratio (SNR) of 30 dB.

Note that the distributed methods were run on a single computer for illustration purposes. In this case, the workers are independent processes.

4.5.1. Compared methods

The estimation performance of the proposed algorithm has been compared to those of several unmixing methods from the literature. Note that only the computation time associated with the Algo. 8 and its synchronous version (both implemented in Julia [Bez+17]) can lead to a consistent comparison in this experiment. Indeed, some of the other unmixing methods have been implemented in MATLAB by their respective authors. In the following lines, implementation details specific to each of these methods are given.

1. VCA/FCLS: the endmembers are first extracted on each image using the vertex component analysis (VCA) [NB05b], which requires pure pixels to be present. The abundances are then estimated for each pixel by solving a fully constrained least squares problem (FCLS) using the ADMM proposed in [BF10];

2. SISAL/FCLS: the endmembers are extracted on each image by the simplex identification via split augmented Lagrangian (SISAL) [Bio09], and the abundances are estimated for each pixel by FCLS. The tolerance for the stopping rule has been set to 10^{-4};

3. Proposed method (referred to as ASYNC): the endmembers have been initialized with the signatures obtained by VCA on the first image of the sequence, and the abundances have been initialized by FCLS. The synchronous and asynchronous algorithms have been stopped when the relative decrease of the objective function between two consecutive iterations is lower than 10^{-4}, with a maximum of 100 and 500 iterations respectively. Its synchronous counterpart is referred to as SYNC;

4. DAVIS [Dav16]: this asynchronous algorithm only differs from the previous algorithm, in that no relaxation step is considered, and in the expression of the descent stepsize used to ensure the algorithm convergence. To ensure a fair

comparison, it has been run in the same setting as the proposed asynchronous method;

5. DSPLR [TRB17]: the DSPLR algorithm has been run with the stopping criterion proposed in [TRB17] (set to $\varepsilon = 10^{-4}$), for a maximum of 100 iterations. The same initialization as the two previous distributed algorithms has been used.

The estimation performance reported in Table 4.1 are evaluated in terms of the performance measures introduced in the preliminary chapter (see section I.4).

4.5.2. Results in absence of variability (LMM)

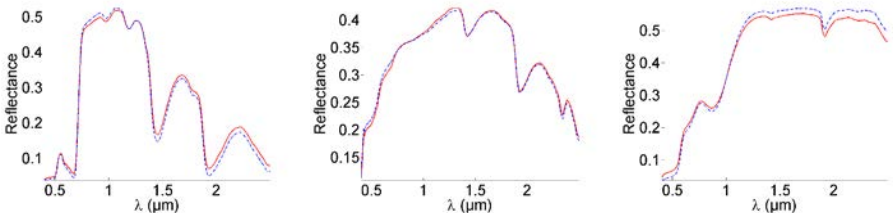
- **Endmember estimation:** the proposed asynchronous algorithm leads to competitive endmember estimation on the three synthetic datasets (in terms of aSAM and RE), notably in comparison with its synchronous counterpart. We can note that the DSPLR algorithm yields interesting estimation results for $R = 3$, which however significantly degrades as R increases. This partly results from the matrix inversions involved in the update steps of [TRB17], which remains relatively sensitive to the conditioning of the involved matrices, and consequently to the choice of the regularization parameter of the augmented Lagrangian. The endmembers extracted from the dataset composed of $R = 3$ endmembers are reported in Fig. 4.3.

- **Abundance estimation:** the synchronous PALM algorithm leads to the best abundance estimation results, even in the absence of any additional regularization on the spatial distribution of the abundances. In this respect, we can note that the performance of PALM and its asynchronous version are relatively similar, which consistently outperform the other unmixing methods. The abundance maps recovered from the dataset composed of $R = 3$ endmembers are reported in Fig. 4.4 for illustration.

- **Overall performance:** the performance measures reported in Table 4.1 show that the proposed distributed algorithm yields competitive estimation results, especially in terms of the required computational time when compared to its synchronous counterpart. To be more explicit, the evolution of the objective as a function of the computation time shows the interest of the allowed asynchronicity to speed up the unmixing task, as illustrated in Fig. 4.5 (the computation time required by Algo. 8 is almost 4 times lower than the one of its synchronous counterpart).

Table 4.1.: Simulation results on synthetic data without variability (GMSE($\mathbf{A}$)$\times 10^{-3}$, RE $\times 10^{-4}$).

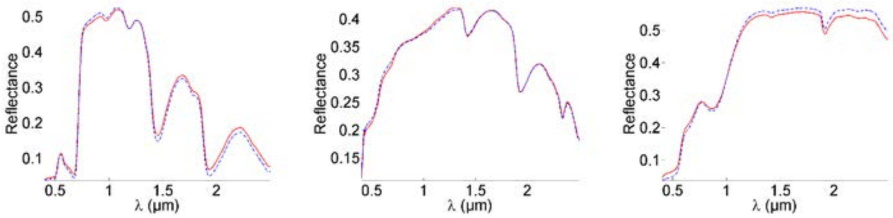
	Algorithm		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	RE	aSAM($\mathbf{Y}$) (°)	time (s)
$R = 3$	VCA/FCLS	[NB05b]	1.82	1.27	0.64	1.45	**1**
	SISAL/FCLS	[Bio09]	1.55	0.94	0.62	1.43	2
	DSPLR	[TRB17]	0.84	2.76	**0.59**	**1.41**	139
	DAVIS	[Dav16]	1.44	0.92	0.63	1.45	10
	SYNC		**0.76**	**0.33**	0.60	**1.41**	197
	ASYNC		0.85	0.38	0.60	**1.41**	101
$R = 6$	VCA/FCLS	[NB05b]	2.55	1.08	1.11	1.64	**1**
	SISAL/FCLS	[Bio09]	1.65	0.50	**0.91**	1.53	2.5
	DSPLR	[TRB17]	3.64	4.65	7.73	**1.45**	116
	DAVIS	[Dav16]	1.87	1.22	0.96	1.58	45.3
	SYNC		**0.63**	**0.28**	**0.78**	**1.45**	462
	ASYNC		1.09	0.59	0.81	1.48	46
$R = 9$	VCA/FCLS	[NB05b]	3.07	2.59	6.75	2.37	**2**
	SISAL/FCLS	[Bio09]	2.17	1.77	5.11	2.14	4
	DSPLR	[TRB17]	8.52	6.53	1.48	**1.56**	153
	DAVIS	[Dav16]	1.57	1.27	1.98	1.69	84
	SYNC		**0.87**	**0.40**	**1.50**	1.57	762
	ASYNC		0.88	0.54	1.52	1.58	170



(a) Material 1 (VCA)

(b) Material 2 (VCA)

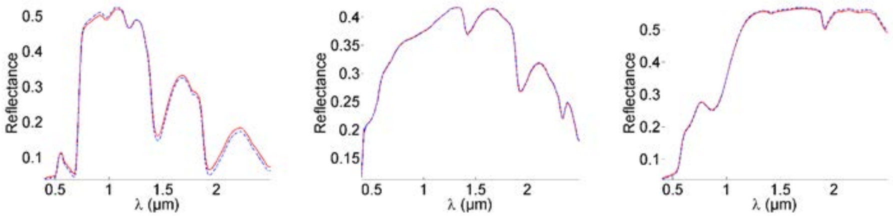
(c) Material 3 (VCA)



(d) Material 1 (SISAL)

(e) Material 2 (SISAL)

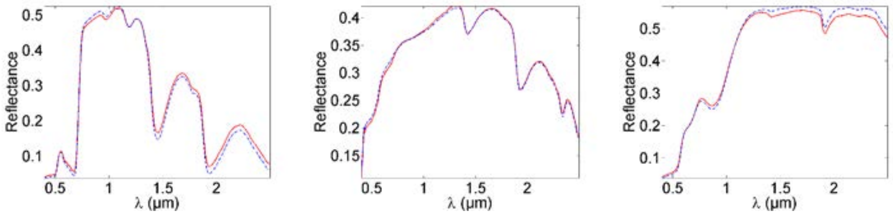
(f) Material 3 (SISAL)



(g) Material 1 (DSPLR)

(h) Material 2 (DSPLR)

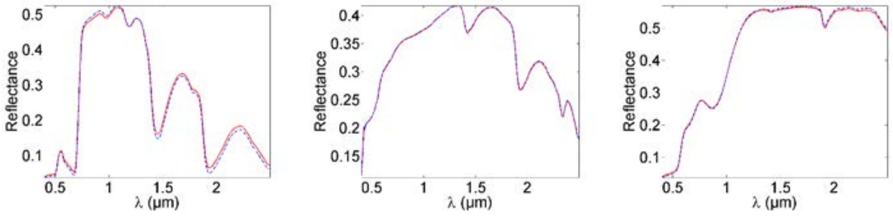
(i) Material 3 (DSPLR)



(j) Material 1 (DAVIS)

(k) Material 2 (DAVIS)

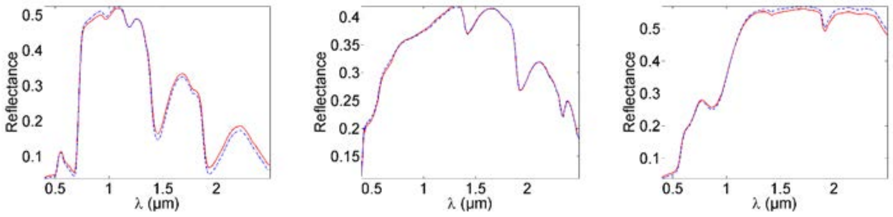
(l) Material 3 (DAVIS)



(m) Material 1 (SYNC)

(n) Material 2 (SYNC)

(o) Material 3 (SYNC)

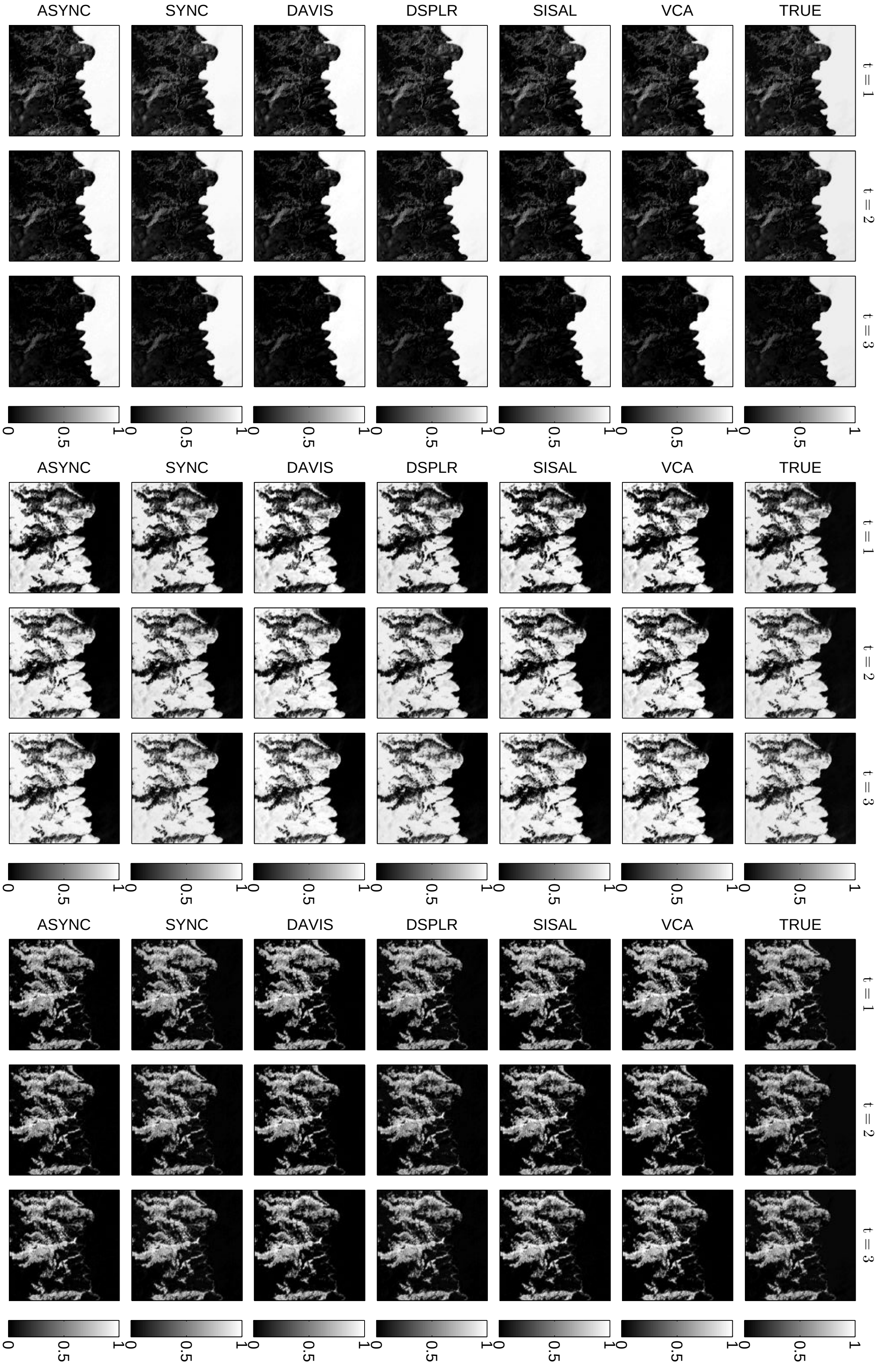


(p) Material 1 (ASYNC)

(q) Material 2 (ASYNC)

(r) Material 3 (ASYNC)

Figure 4.3.: Endmembers (red lines) recovered by the different methods from the synthetic dataset composed of $R = 3$ endmembers. The true endmembers are given in each figure in blue dotted lines for comparison.

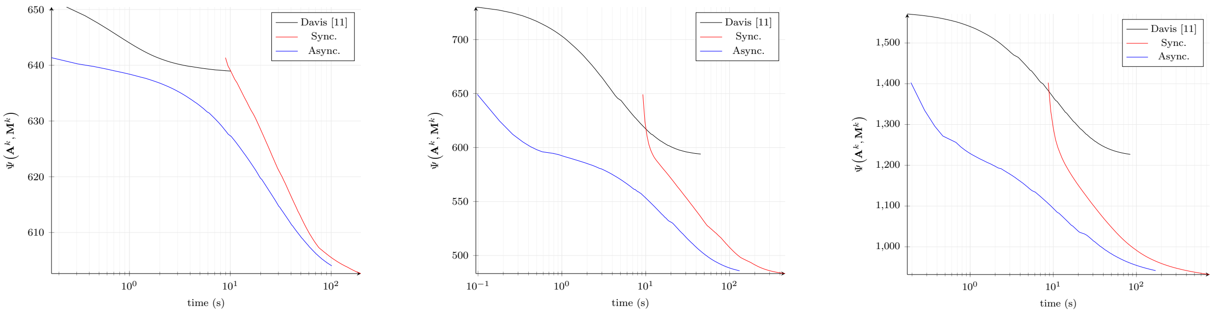


(a) Material 1

(b) Material 2

(c) Material 3

Figure 4.4.: Abundance maps recovered by the different unmixing method from the synthetic dataset composed of $R = 3$ endmembers.

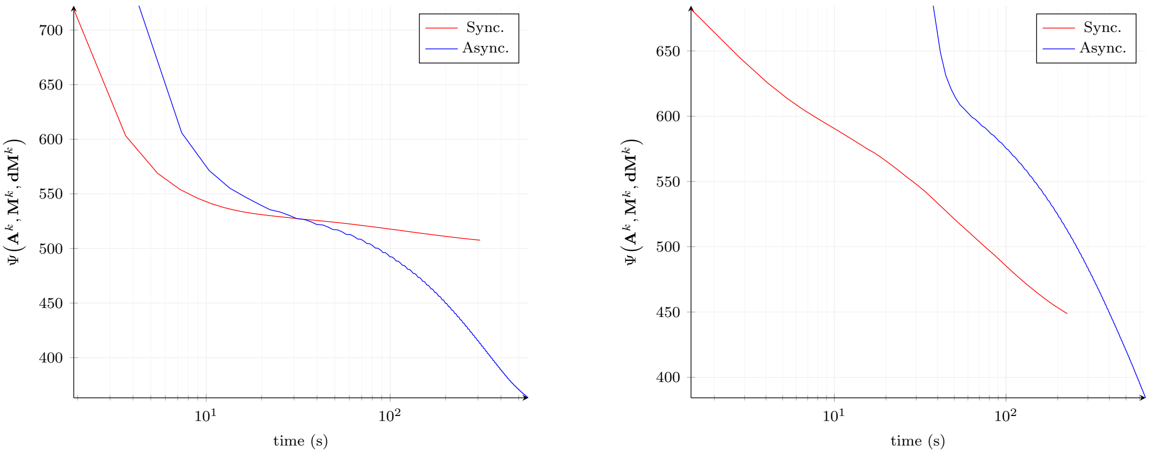


(a) $R = 3$ (b) $R = 6$ (c) $R = 9$

Figure 4.5.: Evolution of the objective function for the synthetic datasets, obtained for Algo. 8 and its synchronous version until convergence (LMM).

4.5.3. Results in presence of variability (PLMM)

In this preliminary experiment, only the VCA/FCLS, SYNC and ASYNC algorithms are compared to illustrate the limits of the distributed unmixing strategy investigated in this chapter.



(a) $R = 3$

(b) $R = 6$

Figure 4.6.: Evolution of the objective function for the synthetic datasets, obtained for Algo. 8 and its synchronous version until convergence (PLMM).

Table 4.2.: Simulation results on synthetic data with variability (GMSE($\mathbf{A}$)$\times 10^{-3}$, GMSE($\mathbf{dM}$)$\times 10^{-5}$, RE $\times 10^{-5}$).

	Algorithm		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	aSAM($\mathbf{Y}$) (°)	time (s)
$R = 3$	VCA/FCLS	[NB05b]	**1.19**	1.40	/	12.72	1.38	**0.94**
	SYNC		**1.19**	**0.80**	**0.55**	5.00	1.26	308
	ASYNC		1.33	0.98	0.97	**3.58**	**1.05**	563
$R = 6$	VCA/FCLS	[NB05b]	2.33	1.70	/	10.67	1.48	**0.80**
	SYNC		**0.81**	**0.82**	**1.09**	7.25	1.36	227
	ASYNC		1.12	1.22	1.26	**6.20**	**1.24**	650

Overall performance: the performance measures reported in Table 4.2 and the evolution of the objective function (Fig. 4.6) show that the proposed distributed algorithm does not yield the expected acceleration in terms of computation time. On the one hand, the SYNC and ASYNC algorithms converge to points whose cost can be significantly different. On the other hand, in comparison with the situation described in the model 4.4, a larger amount of data has to be communicated from a worker node to the master process. This observation, coupled with the increase in the amount of operations performed by the master node to update all the parameters of interest in this context, partly explains the limits of the distributed unmixing approach reported in Fig. 4.6. Parameter updates allowing different network configurations (i.e., as proposed in [PR15; BHI16]) and implementations improvement (data distribution between the workers) should allow a notable acceleration in comparison with the currently proposed method. Given these observations, only the results associated with the model 4.4 are reported for the experiments with real data in the following paragraphs.

4.6. Experiments with real data

In practice, distributed unmixing procedures are of a particular interest when considering the unmixing large HS images, or of a sequence of HS images acquired by possibly different sensors at different time instants [HCJ16; YZP17]. The unmixing of two large real HS images is first proposed, whereas the latter application to MTHS images motivates the last example addressed in the present section.

4.6.1. Description of the datasets

Houston dataset (single HS image). The first dataset considered hereafter has been acquired over the campus of the University of Houston, Texas, USA, in June 2012 [Deb+14]. The 152×108 scene of interest is composed of 144 bands acquired in the wavelength range from $380\,\mathrm{nm}$ to $1050\,\mathrm{nm}$. An equal number of pixels has been assigned to each of the 3 workers used in this experiment.

Cuprite dataset (single HS image). The second dataset considered in this work consists in a 190×250 subset extracted from the extensively studied Cuprite dataset. In this case, reference abundance maps are available from the literature (see for instance [NB05b; MQ07]. After removing water-absorption and low SNR bands, 189 out of the 224 spectral bands initially available were exploited in the subsequent unmixing procedure. The data have been unmixing with $R = 10$ based on prior studies conducted on this dataset [NB05b; MQ07]. The data have been unmixing with $R = 4$ endmembers as in chapter 1. The same number of pixels has been assigned to each of the 3 workers used in this experiment.

Mud lake dataset (MTHS images). We finally consider a real sequence of AVIRIS HS images acquired between 2014 and 2015 over the Mud Lake, located in the Lake Tahoe region (California, United States of America)[2]. The 100×100 scene of interest is in part composed of a lake and a nearby field. The images have been unmixed with $R = 3$ endmembers based on results obtained from the prior studies conducted on these data in chapters 2 and 3, further confirmed by the the results of the noise-whitened eigengap algorithm (NWEGA) [Hal+16b] reported in Table 4.3. After removing the water absorption bands, 173 out of the 224 available spectral bands were finally exploited. The pixels corresponding to 2 out of the 6 images have been assigned to each of the 3 available workers.

Table 4.3.: Endmember number R estimated by NWEGA [Hal+16b] on each image of the Mud lake dataset.

	04/10/2014	06/02/2014	09/19/2014	11/17/2014	04/29/2015	10/13/2015
NWEGA	3	3	3	4	3	4

---

[2] The images from which the interest of interest is extracted 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 (f) 10/13/2015

Figure 4.7.: Mud lake dataset used in the MTHS experiment, given with their respective acquisition date. The area delineated in red in Fig. 4.7(e) highlights a region known to contain outliers (this observation results from the analysis led on this dataset in chapter 2).

4.6.2. Results

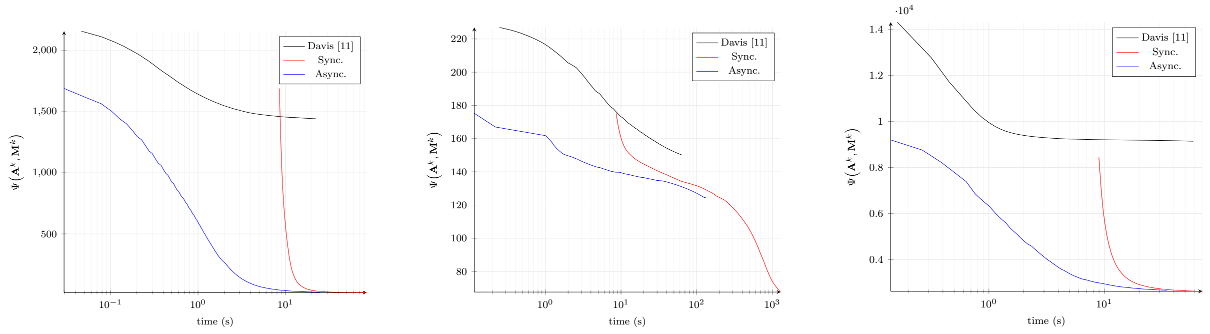
Given the absence of ground truth for the different datasets (except the indications available in the literature for the Cuprite scene [NB05b; MQ07]), the estimation results obtained by the proposed algorithms are compared to the other unmixing procedures in terms of the RE and the aSAM (see Table 4.4) introduced in section I.4 of the preliminary chapter. The consistency of the estimated abundance maps, reported in Figs. 4.9 to 4.13, is also considered when analyzing the different results.

Houston. The distributed algorithms yield abundance maps in agreement with the VCA/FCLS and SISAL algorithms (see Fig. 4.10). We can note that the algorithms SYNC, ASYNC and DSPLR provide a more contrasted abundance map for the concrete than VCA/FCLS, SISAL/FCLS and DAVIS.

Cuprite. Except for the DSPLR algorithm, whose scale indeterminacy leads to results somewhat harder to interpret for this dataset, the results obtained by the different methods are relatively similar for the different datasets, be it in terms of the estimated abundance maps or the recovered endmembers.

Mud lake. The algorithms SYNC, DAVIS [Dav16] and ASYNC lead to particularly convincing abundance maps, in the sense that the abundances of the different materials (containing soil, water and vegetation) are consistently estimated (see Figs. 4.11 to 4.13). At $t = 5$, VCA/FCLS and SISAL, which have been applied individually to each image of the sequence, appear to be particularly sensitive to the presence of outliers in the area delineated in red in Fig. 4.7(e) (see chapter 2). This observation is further confirmed by the endmembers reported in Fig. 4.14, whose amplitude is significantly greater than 1. This sensitivity notably results from the fact that each scene has been analyzed independently from the others in this specific context (note that the results would have been worse if these methods were applied to all the images at once).

Global reconstruction performance. The performance measures reported for the different datasets in Table 4.4 confirm the interest of the PALM algorithm and its asynchronous variant for unmixing applications. The asynchronous variant can be observed to lead to a notable reduction of the computation time (see also Fig. 4.8), while allowing a reconstruction performance similar to the classical PALM algorithm to be obtained.



(a) Houston (b) Cuprite (c) Mud lake

Figure 4.8.: Evolution of the objective function for the synthetic datasets, obtained for DAVIS [Dav16], Algo. 8 and its synchronous version until convergence (model (4.4)).

Table 4.4.: Simulation results on real data (RE $\times 10^{-4}$).

	Algorithm		RE	aSAM($\mathbf{Y}$) (°)	time (s)
Cuprite	VCA/FCLS	[NB05b]	0.51	0.96	**2**
	SISAL/FCLS	[Bio09]	0.47	0.92	6
	DSPLR	[TRB17]	1.25	1.42	20.2
	DAVIS	[Dav16]	0.33	0.79	64.0
	SYNC		**0.15**	**0.55**	1290
	ASYNC		0.30	0.77	134
Houston	VCA/FCLS	[NB05b]	22.5	3.31	**0.1**
	SISAL/FCLS	[Bio09]	21.3	2.01	0.6
	DSPLR	[TRB17]	**0.13**	**0.99**	51.5
	DAVIS	[Dav16]	14.9	2.44	22.3
	SYNC		0.21	1.14	84.6
	ASYNC		0.24	1.17	24.9
Mud lake	VCA/FCLS	[NB05b]	23.7	13.23	**1**
	SISAL/FCLS	[Bio09]	**1.65**	**3.09**	2
	DSPLR	[TRB17]	1.93	10.9	99.6
	DAVIS	[Dav16]	17.61	6.27	58.9
	SYNC		5.05	5.88	70.4
	ASYNC		5.13	5.88	35.0

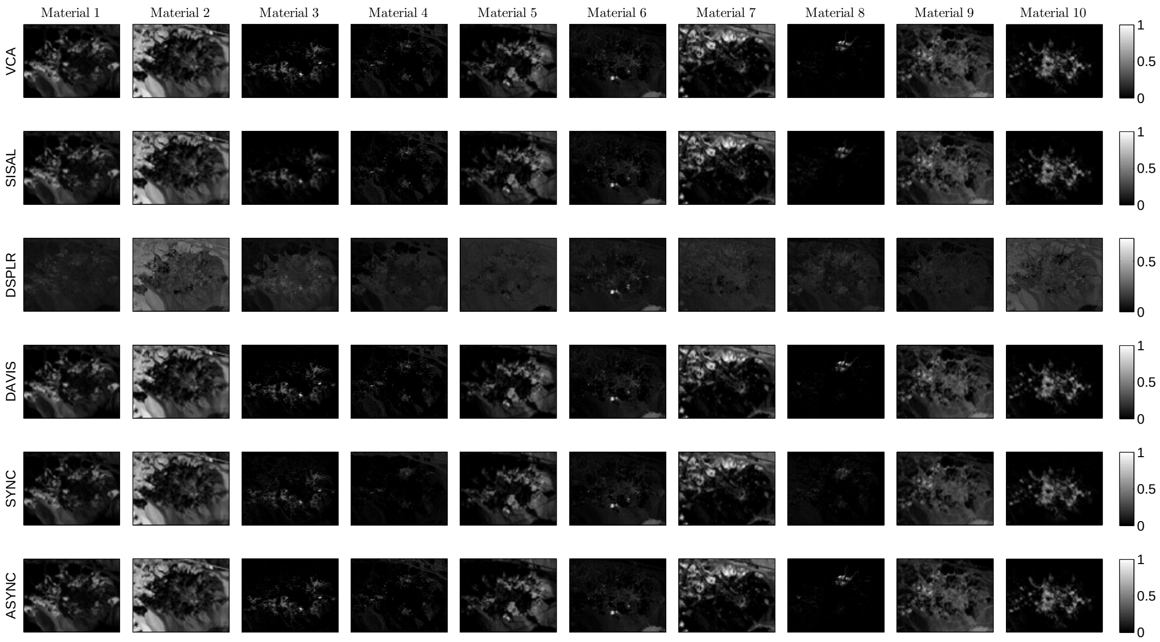


Figure 4.9.: Abundance maps recovered by the different methods (in each row) for the Cuprite dataset.

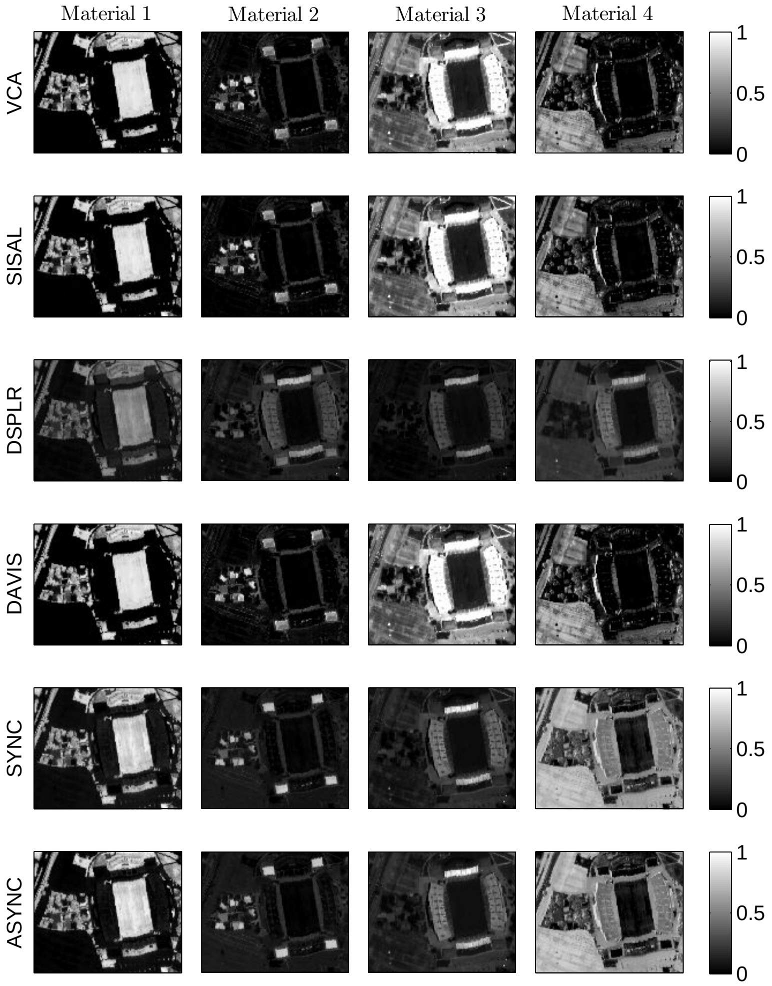


Figure 4.10.: Abundance maps recovered by the different methods (in each row) for the Houston dataset.

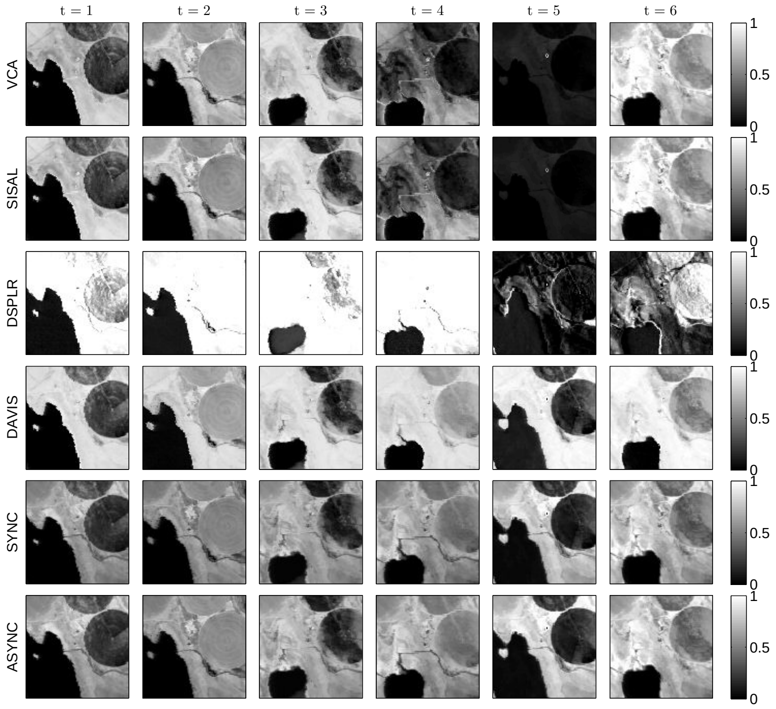


Figure 4.11.: Soil abundance map recovered by the different methods (in each row) at each time instant (given in column) for the experiment on the Mud lake dataset [the different rows correspond to VCA/FCLS, SISAL/FCLS, DSPLR [TRB17], ASYNC [Dav16], SYNC and ASYNC methods].

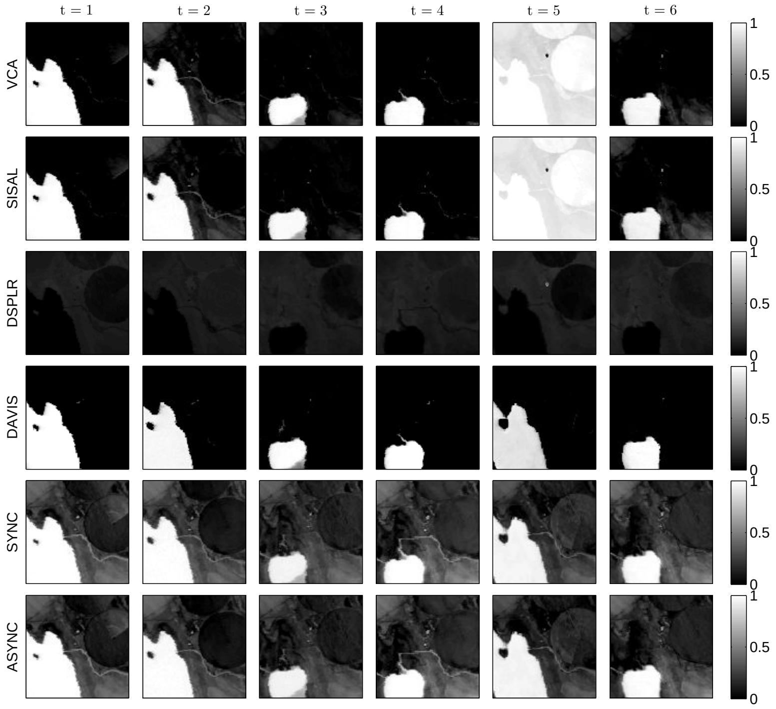
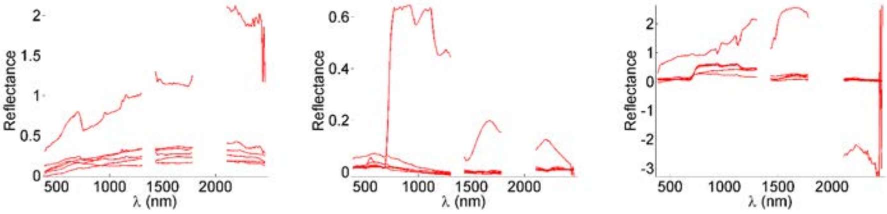
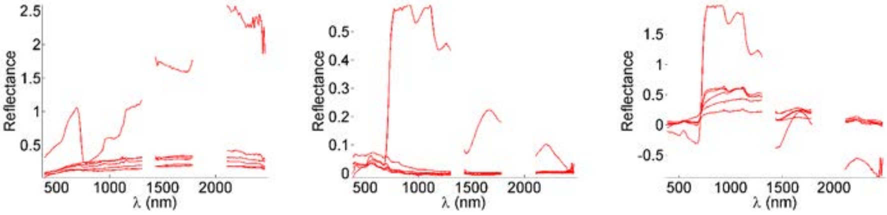


Figure 4.12.: Water abundance map recovered by the different methods (in each row) at each time instant (given in column) for the experiment on the Mud lake dataset [the different rows correspond to VCA/FCLS, SISAL/FCLS, DSPLR [TRB17], ASYNC [Dav16], SYNC and ASYNC methods].

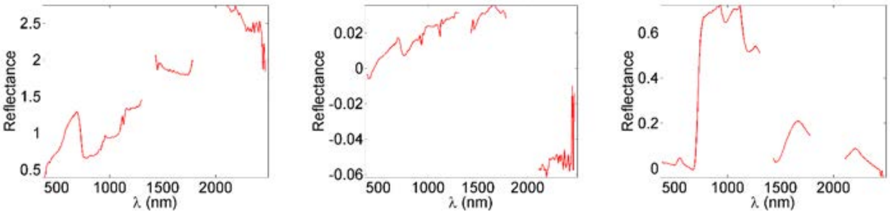
Figure 4.13.: Vegetation abundance map recovered by the different methods (in each row) at each time instant (given in column) for the experiment on the Mud lake dataset [the different rows correspond to VCA/FCLS, SISAL/FCLS, DSPLR [TRB17], ASYNC [Dav16], SYNC and ASYNC methods].



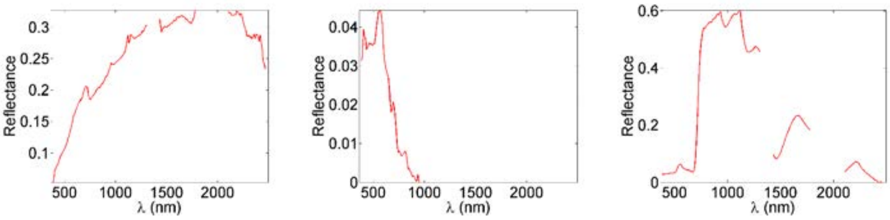
(a) Soil (VCA) (b) Water (VCA) (c) Veg. (VCA)



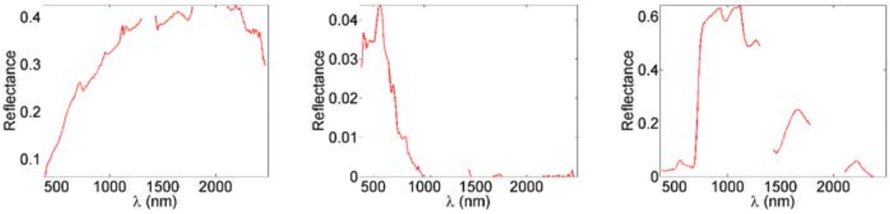
(d) Soil (SISAL) (e) Water (SISAL) (f) Veg. (SISAL)



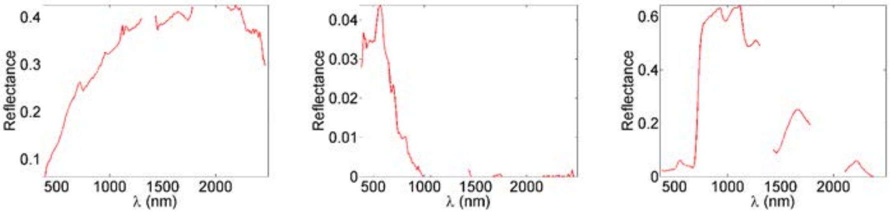
(g) Soil (DSPLR) (h) Water (DSPLR) (i) Veg. (DSPLR)



(j) Soil (DAVIS) (k) Water (DAVIS) (l) Veg. (DAVIS)



(m) Soil (SYNC) (n) Water (SYNC) (o) Veg. (SYNC)



(p) Soil (ASYNC) (q) Water (ASYNC) (r) Veg. (ASYNC)

Figure 4.14.: Endmembers ($\mathbf{m}_r$, red lines) recovered by the different methods from the real dataset depicted in Fig. 4.7.

4.7. Conclusion and future work

This chapter was focused on the design of a partially asynchronous algorithm based on [Can+16; Cha+16; Dav16], which proves particularly adapted to address large scale matrix factorization problems. Under relatively standard conditions, the proposed approach inherits from the convergence guarantees studied in [Can+16], and from the traditional PALM algorithm [BST13; CPR16] for its synchronous counterpart. Evaluated on a hyperspectral unmixing problem, the proposed approach provided competitive estimation results in absence of variability, while significantly reducing the computation time to reach convergence.

Future research perspectives include the extension to different network topology as can be used in [PR15; BHI16], or the use of variable metrics as described in [RCP14; CPR14; CPR16; FGP15].

4.8. Conclusion (in French)

Ce chapitre a permis d'étudier un algorithme de démélange partiellement asynchrone basé sur de récentes contributions en analyse non-convexe [Can+16; Cha+16; Dav16], particulièrement adapté à la résolution de problèmes de factorisation matricielle en grande dimension. Sous un ensemble d'hypothèses standards, l'approche proposée hérite des garanties de convergences de l'algorithme proposé par [Can+16], et de celles de l'algorithme PALM [BST13; CPR16] pour sa version synchrone. Évalué dans le cadre du démélange hyperspectral, l'algorithme asynchrone présente des résultats d'estimation satisfaisants en l'absence de variabilité par rapport à une version synchrone du même algorithme, tout en réduisant de façon significative le temps de calcul nécessaire pour parvenir à convergence.

L'ajout de métriques variables, telles que décrites dans [RCP14; CPR14; CPR16; FGP15], ou l'utilisation d'une topologie différente du réseau de calcul [BHI16; PR15] comptent parmi les perspectives de recherche qu'il conviendrait d'approfondir.

Conclusions and perspectives

Context

Acquired in hundreds of contiguous spectral bands, hyperspectral images present a comparatively more limited spatial resolution in specific remote sensing applications. In this case, the observations are represented as the mixture of several signatures corresponding to distinct materials, which have to be estimated. Though each observed material is traditionally assumed to be characterized by a single spectrum, varying acquisition conditions can induce a notable variability in the extracted signatures, both locally and temporally (depending on the acquisition date). Unless specifically accounted for, this phenomenon can result in severe estimation errors, particularly in the case of unsupervised unmixing procedures. An additional computational hurdle arises from the possibly significant size of the data, especially when analyzing multitemporal hyperspectral images.

To address these issues, the present manuscript has first investigated a new explicit mixture model to represent spatial endmember variability, later extended to account for temporal variability. Based on this model, different algorithmic approaches have been considered to efficiently solve the resulting high dimensional estimation problems, involving single or multitemporal hyperspectral images. The study conducted in this manuscript has allowed the following conclusions to be drawn.

Conclusions

Chapter 1 has introduced a new explicit mixture model inspired from the total least squares problem, referred to as perturbed linear mixture model (PLMM), to account for spatial endmember variability within a single hyperspectral (HS) image. The parameters of the proposed model have been estimated by a first algorithm based on the ADMM method, and a second algorithm based on a proximal alternating linearized minimization (PALM) [BST13; CPR16]. The results obtained on real data have notably exemplified the relevance of the proposed model to handle scenes of moderate size, *a priori* affected by an appreciable level of variability. In practice, the proposed approach has provided a relevant source of information on the spatial distribution of the variability within the scene under study. In comparison to the ELMM described in [Dru+16, (4)], the PLMM has additionally allowed information on the variability level affecting each spectral band to be captured. Determining which spectral bands are affected by spectral variability can be of interest to the practitioner, e.g., to characterize the source of the observed variability, or to determine whether spectral bands with a low SNR have not been removed prior to the unmixing process.

Chapter 2 has investigated a robust variant of the model considered in chapter 1 to represent temporal endmember variability and abrupt spectral changes commonly observed in multitemporal HS (MTHS) images. Starting from the results obtained on a reference HS image, the observations are likely to be composed of the same materials, except in a few pixels of the time series where new materials may have appeared. In this context, the proposed approach consists in interpreting smooth spectral variations in terms of endmember variability, while allowing the influence of abrupt spectral changes induced by outliers (or new endmembers) to be mitigated. In practice, a variant of the PLMM has allowed information redundancy to be exploited, while reducing the impact of significant spectral changes induced by outliers. A Markov chain Monte-Carlo method has been proposed to precisely assess the performance of the model on synthetic and real data, yielding encouraging results in the latter case. More precisely, the PLMM has provided a fine characterization of the evolution of the materials observed in the reference image, and has captured outliers in pixels known to be corrupted by a sensor defect. Moreover, the joint analysis of several HS images has been shown to significantly improve the unmixing results when compared to methods analyzing the images independently.

An online unmixing algorithm has been considered in chapter 3 to address large datasets which cannot be loaded in memory at once. In contrast with the approach described in the previous paragraph, the data are progressively incorporated into the unmixing process, thus reducing the memory requirements. This type of approach is particularly relevant to analyze MTHS images, as those considered in chapter 2. Experiments conducted on synthetic data have illustrated the competitiveness of the proposed online unmixing algorithm with respect to two reference batch algorithms (i.e., when the data can be exploited by a batch or an online algorithm). In agreement with the conclusions drawn in chapter 2, the experiments conducted on both synthetic and real data have shown the relevance of exploiting several images to improve the unmixing performance. In addition, the energy of the instantaneous endmember variability has provided a complementary information to identify at which time instant a given endmember has deviated the most from its average spectral signature.

In order to provide a complementary approach to the study conducted in chapter 3, chapter 4 has illustrated the interest of the asynchronicity allowed by recent distributed algorithms to decrease the computation time required to reach convergence. The relevance of asynchronicity for unmixing applications is evaluated in comparison with a synchronous distributed version of the same algorithm. Evaluated on the linear mixture model (LMM), the asynchronous unmixing procedure has led to a notable acceleration, at the price of a slightly degraded precision in the parameter estimation. The current limitations of the distributed strategy used in this work has been illustrated on synthetic data when using the PLMM. Finally, several experiments conducted on real datasets have illustrated the performance of this approach.

Perspectives and future work

The present study has raised several research perspectives summarized in the following lines.

Variability modeling

From a modeling point of view, the PLMM considered in chapter 1 does not address the semantic ambiguity resulting from the physical nature of the observed variability. Whereas the ELMM introduced in [Dru+16] has a clear interpretation in terms of illumination variations, the PLMM does not lead to a clear distinction between the potential sources of variability, e.g., illuminations variations or non-linearities. As such, the design of physically inspired algorithms, which possibly preserve a distinction between the different sources of variability, can be further investigated to capture as much information as possible with respect to the observed materials.

From a different perspective, the use of different priors on the variability structure may reduce the indeterminacy inherent to the model proposed in chapter 1, i.e., between the endmembers and the captured variability. In practice, endmember variability can be expected to affect only a few pixels within the image under study, e.g., when resulting from non-linearities. In this case, a spatial sparsity-promoting prior defined as

$$\Upsilon(\mathbf{dM}) = \sum_{r,n} \gamma_{r,n} \|\mathbf{dm}_{r,n}\|_2, \quad \gamma_{r,n} > 0$$

can be easily addressed by the PALM algorithm described in section 1.4.2. The only modification appears in the proximal operator involved in the update equation 1.23, which can be easily computed by using (1.24) and [CP11], [Yu13, Theorem 4]. The problem then consists in designing appropriate strategies to automatically select the value of the regularization parameters $\gamma_{r,n}$, which is currently under study. A complementary study can be conducted on real data to assess the relevance of applying sparsity promoting priors to endmember variability in a transformed domain [Rap+14], e.g., by defining

$$\Upsilon(\mathbf{dM}) = \sum_{r=1}^{R} \big\|\boldsymbol{\Lambda} \odot \big([\mathbf{dm}_{r,1}, \dots, \mathbf{dm}_{r,N}]\boldsymbol{\Phi}^{\mathrm{T}}\big)\big\|_1, \quad \boldsymbol{\Lambda} \succeq \mathbf{0}_{R,K}, \quad , \boldsymbol{\Phi} \in \mathbb{R}^{K \times N}, \quad \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \mathbf{I}_K$$

where $\boldsymbol{\Phi}$ is an appropriate transform domain, and $\boldsymbol{\Lambda}$ denote the associated regularization parameters. On the other hand, the concept of morphological diversity [Bob+07] would deserve to be explored to further discriminate the contribution of endmember variability from the signal of interest.

Computational aspects

Hyperparameter selection. The unmixing problems formulated in chapters 1 and 3 depend on several hyperparameters whose value has been chosen by testing several values on a predefined grid. In practice, the use of automatic estimation strategies within the algorithms developed in this manuscript remains to be studied. For instance, the applicability of approaches based on the Stein's unbiased risk estimate [Ste81; Del+14] can be first explored. Alternatively, the methods developed in [PBF15; Hal+17] assign appropriate priors to the hyperparameters (typically gamma priors), which are then estimated along the other parameters of interest. The resulting non-convex problem is then solved by an alternating optimization algorithm, such as a block coordinate descent [Wri15]. Another possibility may consist in adopting a two step strategy as proposed in [Fre+17].

Computational improvements.

- Variable metrics [CPR14; CPR16; FGP15] can be used to extend the algorithms proposed in chapters 1 and 4. This extension is expected to allow a more efficient estimation of the parameters (in terms of convergence speed) when compared to the standard PALM [BST13] used in this manuscript (see [RCP14] for instance). In particular, this acceleration may prove beneficial to mitigate the limitations observed with the asynchronous distributed unmixing algorithm studied in chapter 4.
- Relaxations to the Ising field used in chapter 2 could be contemplated to estimate the parameters of the Bayesian model developed in section (2.4). This modification would notably allow the development of an online unmixing algorithm as the one discussed in chapter 3, thus reducing the memory requirements when compared to the initial formulation of the problem. In this perspective, replacing the discrete labels by continuous latent variables as proposed in [MSB03; BCK14; BF16] can be investigated to describe the outliers' support.
- Complementary simulations on larger datasets (possibly running on several machines) would allow the interest of the distributed approach considered in chapter 4 to be better characterized in the context of hyperspectral unmixing. In particular, a finer analysis of the computational bottlenecks arising in the context of the PLMM is needed to overcome the algorithmic limitations currently observed in chapter 4. This study will be essential to ensure the characterization of spatio-temporal variability within the numerous multi-temporal images that will be regularly acquired in the coming years (by the European Space agency[1] for instance). In addition, several distributed primal-dual algorithms from the literature [PR15; BHI16] can be compared to assess the interest of several distributed optimization architectures for remote sensing applications.

Application-oriented developments

The model introduced in chaper 1 can be used in different contexts, e.g., in medical imagery [Cav+17]. More precisely, the work presented in [Cav+17] for positron emission tomography (PET) resorts to a PLMM-based unmixing approach to model the time-activity curves associated with the concentration evolution of a radiotracer in the body. Besides, the use of a spatially varying dictionary to analyze astronomical data [Rap+14; CBR15; CB17] might provide a complementary information with respect to the spatial evolution of the sources of interest, which have to be estimated.

In addition, the relevance of the approaches proposed in chapters 2 and 3 could be evaluated in the context of change detection problems involving images acquired with possibly different imaging modalities [Pre+16; YZP17; Fer+17]. For this application, the resulting unmixing problem can be formulated as a joint fusion and unmixing problem [Wei+16], where the LMM underpinning the approach described in [Wei+16] is replaced by the PLMM.

---

[1] `http://www.esa.int/Our_Activities/Observing_the_Earth/Proba-1/Going_hyperspectral`

Conclusions et perspectives (in French)

Contexte

Acquises dans plusieurs centaines de bandes spectrales contiguës avec une résolution spectrale fine, les images hyperspectrales présentent une résolution spatiale moins fine pour certaines applications en télédétection. Dans ce cas, les acquisitions sont représentées par un mélange de plusieurs signatures caractéristiques des différents matériaux observés (spectres purs) qu'il convient d'estimer. Bien qu'il soit d'usage de supposer qu'une unique signature permet de caractériser un matériau donné, la variation des conditions d'acquisition peut induire une variabilité notable des signatures extraites. À défaut d'être prise en compte, cette source de variabilité peut induire d'importantes erreurs d'estimation, notamment dans le cas de procédures de démélange non-supervisées. À cette problématique viennent s'ajouter différentes limitations calculatoires liées à la dimension des données à exploiter, notamment dans le cas d'images hyperspectrales multitemporelles.

Pour aborder ces deux problématiques, un nouveau modèle de mélange a tout d'abord été introduit pour représenter explicitement la variabilité spatiale et temporelle des acquisitions. Différentes approches algorithmiques ont ensuite été proposées pour aborder les problèmes d'estimation résultant, dans le cas d'images hyperspectrales mono- et multi-temporelles. L'étude présentée dans ce manuscrit, et résumée ci-après, a permis de formuler les conclusions suivantes.

Conclusions

Le premier chapitre a permis d'introduire un modèle de mélange visant à représenter explicitement la variabilité spatiale des spectres purs au sein d'une image hyperspectrale. Inspiré du problème des moindres carrés totaux [GL80], les paramètres de ce modèle ont été estimés à l'aide d'un premier algorithme basé sur l'algorithme ADMM, et d'un second basé sur l'algorithme PALM. Les résultats obtenus sur données réelles ont illustré la pertinence de l'approche proposée pour rendre compte de la variabilité observée dans des images de taille modérée, affectées par un niveau appréciable de variabilité. Par ailleurs, cette approche a permis de cartographier la distribution spatiale de la variabilité au sein de la scène, ainsi que d'identifier les bandes spectrales principalement affectées par la variabilité. Ces informations peuvent permettre de caractériser les causes de la variabilité spectrale observée, et de déterminer si certaines bandes à faible rapport signal-sur-bruit n'ont pas été exclues des données traitées.

Le deuxième chapitre a étudié une variante du modèle considéré au chapitre 1 pour représenter la variabilité temporelle des spectres purs dans une série d'images hyperspectrales. Partant d'une scène de référence, les autres images de la série sont consituées en majeure partie des mêmes matériaux, hormis en un nombre limité de pixels en lesquels des nouveaux matériaux peuvent être observés. Dans ce contexte, l'approche proposée consiste à interpréter les variations modérées des spectres purs en termes de variabilité, et à considérer les changements abrupts comme des données aberrantes. Un algorithme de Monte-Carlo par chaîne de Markov a permis d'estimer les paramètres du modèle, qui a conduit à des résultats encourageants sur données réelles. De fait, ce modèle a permis d'obtenir une caractérisation relativement fine de l'évolution des matériaux présents dans la scène de référence, tout en traduisant la présence de données aberrantes pour des pixels en lesquels un défaut du capteur a été identifié *a priori*. Ce chapitre a notamment attesté la pertinence d'une analyse conjointe de plusieurs images hyperspectrales pour améliorer les résultats du démélange, comparée à une analyse individuelle de chacune des images disponibles.

Une procédure de démélange en ligne a été étudiée au chapitre 3 dans la perspective de traiter des problèmes de grande dimension, pour lesquels l'ensemble des données ne peut être simultanément chargé en mémoire. Contrairement à l'approche considérée au chapitre précédent, les données sont progressivement intégrées au processus de démélange, ce qui permet de réduire le coût mémoire engendré. Ce type d'approche s'avère particulièrement pertinent pour l'analyse d'images hyperspectrales multitemporelles (voir chapitre 2). Des simulations conduites sur données synthétiques et réelles ont permis d'illustrer l'intérêt d'une approche en ligne par rapport à un algorithme traitant l'ensemble des données simultanément (pour des données pouvant être traitées indifféremment par l'une ou l'autre des deux approches). Par ailleurs, les résultats sur données réelles ont montré l'intérêt de considérer plusieurs images pour le démélange, en accord avec les conclusions du chapitre 2.

Enfin, le chapitre 4 a étudié la mise en œuvre d'une procédure de démélange asynchrone, apportant un point de vue complémentaire sur le problème traité au chapitre 3. En particulier, le chapitre 4 a illustré la contribution apportée par l'asynchronie autorisée par de récents algorithmes distribués, dans le but de réduire le temps de calcul requis pour parvenir à convergence. Comparé à une version synchrone du même algorithme pour un problème de démélange linéaire, l'approche considérée a permis d'obtenir une réduction notable du temps de calcul sur données synthétiques et réelles, au prix du'une légère dégradation des performances d'estimation. Les limites de cet algorithme ont par ailleurs été mises en évidence sur données synthétiques lors de l'estimation des paramètres du modèle de mélange en présence de variabiltié spatiale.

Perspectives de recherche

L'étude menée dans ce manuscrit a conduit à proposer plusieurs perspectives de recherche, résumées dans les paragraphes suivants.

Représentation de la variabilité

En ce qui concerne la modélisation de la variabilité, le modèle introduit au chapire 1 ne permet pas de résoudre l'ambiguïté sémantique qui persiste quant à la nature de la variabilité capturée par ce modèle. Tandis que le modèle linéaire étendu proposé dans [Dru+16] présente une interprétation explicite en termes de variation des conditions d'illumination, le modèle introduit au premier chapitre ne permet pas de préserver une nette distinction entre différentes sources de variabilité (e.g., conditions d'illumination et non-linéarités). De ce point de vue, d'autres modèles explicites fondés sur des considérations physiques restent à proposer, afin de caractériser au mieux les matériaux observés.

Dans une toute autre perspective, l'utilisation d'informations *a priori* complémentaires quant à la structure de la variabilité peuvent permettre de pallier l'ambiguïté inhérente au modèle de mélange perturbé du premier chapitre. De fait, la variabilité spatiale n'affecte qu'un nombre de pixels *a priori* restreint lorsqu'elle résulte de non-linéarités. Ces considérations peuvent conduire à utiliser une pénalité spatiale définie par

$$\Upsilon(\mathbf{dM}) = \sum_{r,n} \gamma_{r,n} \|\mathbf{dm}_{r,n}\|_2, \quad \gamma_{r,n} > 0$$

facilement prise en compte dans le cadre de l'algorithme PALM décrit dans la section 1.4.2. La seule modification apportée concerne l'opérateur proximal impliqué dans l'équation 1.23, qui peut être calculé facilement à partir de (1.24) et [CP11], [Yu13, Theorem 4]. La problématique d'intérêt concerne la définition d'une procédure automatique de choix des pondérations $\gamma_{r,n}$, ce qui est actuellement à l'étude. De plus, une étude complémentaire peut être menée afin d'évaluer la pertinence d'une pénalité spatiale dans un domaine transformé [Rap+14], obtenue par exemple en considérant

$$\Upsilon(\mathbf{dM}) = \sum_{r=1}^{R} \|\boldsymbol{\Lambda} \odot \big([\mathbf{dm}_{r,1}, \dots, \mathbf{dm}_{r,N}]\boldsymbol{\Phi}^{\mathrm{T}}\big)\|_1, \quad \boldsymbol{\Lambda} \succeq \mathbf{0}_{R,K}, \quad , \boldsymbol{\Phi} \in \mathbb{R}^{K\times N}, \quad \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \mathbf{I}_K$$

où $\boldsymbol{\Phi}$ constitue un domaine transformé approprié. Par ailleurs, il conviendrait d'étudier le concept de diversité morphologique [Bob+07] dans la perspective de mieux discriminer les sources de variabilité spectrale.

Aspects calculatoires

Sélection des hyperparamètres. Les problèmes de démélange formulés aux chapitres 1 et 3 font intervenir plusieurs hyperparamètres, dont la valeur a été fixée en testant plusieurs valeurs prises sur une grille prédéfinie. La mise en place de procédures de choix automatisées de ces paramètres reste à étudier dans le cadre des algorithmes présentés jusqu'ici. Une approche fondée sur [Ste81; Del+14] pourrait notamment être envisagée. Une alternative peut également consister à assigner une information *a priori* au hyperparamètres, puis à les estimer conjointement avec les paramètres d'intérêts à la manière de [PBF15; Hal+17]. Une autre possibilité serait d'adopter une procédure en deux étapes, telle que celle proposée par les auteurs de [Fre+17].

Améliorations et extensions des algorithmes proposés.

- L'ajout de métriques variables, telles que proposées dans [RCP14; CPR14; CPR16; FGP15] devraient permettre une estimation plus efficace des paramètres du modèle introduit au chapitre 1, notamment en termes du nombre d'itérations nécessaires pour parvenir à convergence. En particulier, une telle accélération pourrait permettre de corriger en partie les limitations de l'algorithme asynchrone proposé au chapitre 4.
- Une relaxation du modèle de champ d'Ising considéré au chapitre 2 pourrait être envisagée dans le but de mettre en œuvre une procédure de démélange en ligne similaire à celle du chapitre 3. Suivant l'exemple de [MSB03; BCK14; BF16], une première approche pourrait consister à remplacer les variables discrètes du modèle hiérachiques par des variables continues afin de décrire le support des données aberrantes.
- Des simulations complémentaires, conduites sur de plus grands jeux de données (notamment sur plusieurs machines distantes), devraient permettre de mieux caractériser les avantages et limites de la procédure de démélange distribuée étudiée au chapitre 4. En particulier, une analyse détaillée du coût calculatoire engendré par chacune des étapes de l'algorithme (lors de son implantation) s'avère essentiel dans la perspective de traiter les nombreuses images hyperpectrales qui seront régulièrement acquises dans les prochaines années (notamment par l'Agence spatiale européenne[1]). Enfin, différents algorithmes distribués primaux-duaux de la littérature [PR15; BHI16] pourront être comparés afin évaluer l'intérêt de différentes architectures d'optimisation distribuées pour des problèmes de démélange.

Applications pratiques

Le modèle du chapitre 1 peut présenter un intérêt dans le cadre d'autres applications, par exemple dans le cadre de la tomographie par émission de positron (TEP) en imagerie médicale [Cav+17]. En particulier, l'étude présentée dans [Cav+17] s'appuie sur un modèle de mélange inspiré de celui introduit dans le premier chapitre pour représenter les courbes d'activité associées à l'évolution de la concentration de radio-traceurs dans le corps. Par ailleurs, l'utilisation d'un modèle de variabilité spatiale devrait permettre d'apporter une infromation complémentaire quant aux sources étudiées en astrophysique ([Rap+14; CBR15; CB17]).

De plus, l'approche proposée au chapitre 2 pourrait être évaluée dans le contexte d'un problème de détection de changement, impliquant des images issues de différentes modalités d'acquisition [Pre+16; YZP17; Fer+17]. Dans ce dernier cas, le problème pourrait être formulé sous la forme d'un problème conjoint de démélange et de fusion de données [Wei+16] basé sur le modèle considéré dans le premier chapitre.

---

[1] `http://www.esa.int/Our_Activities/Observing_the_Earth/Proba-1/Going_hyperspectral`

APPENDIX **A**

Appendix to chapter 1

A.1. Constraints and penalization terms

A.1.1. Abundance penalization: spatial smoothness

The abundance smoothness is expressed in matrix form as

$$\Phi(\mathbf{A}) = \frac{1}{2}\left\|\mathbf{A}\mathbf{H}\right\|_{\mathrm{F}}^{2} \tag{A.1}$$

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

$$\mathbf{H} = \Big[\mathbf{H}_{\leftarrow}, \mathbf{H}_{\rightarrow}, \mathbf{H}_{\uparrow}, \mathbf{H}_{\downarrow}\Big] \in \mathbb{R}^{N\times 4N}.$$

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

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

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

$$\mathbf{H}_{\uparrow} = \Big[\mathbf{0}_{N,W}, \mathbf{H}_u\Big] \quad \text{and} \quad \mathbf{H}_{\downarrow} = \Big[\mathbf{H}_d, \mathbf{0}_{N,W}\Big]$$

with

$$\mathbf{H}_u = \left(\begin{array}{c|cccc} & -1 & 0 & \dots & 0 \\ W & & \ddots & \ddots & \vdots \\ & 1 & & \ddots & 0 \\ \hline & 0 & \ddots & & -1 \\ N-W & \vdots & \ddots & \ddots & \\ & 0 & \dots & 0 & 1 \end{array}\right) \in \mathbb{R}^{N\times(N-W)}, \quad \mathbf{H}_d = -\mathbf{H}_u.$$

The only terms in $\Phi(\mathbf{A})$ related to $\mathbf{a}_n$ are

$$\phi(\mathbf{a}_n) = \frac{1}{2}\underbrace{\left(\sum_{k=0}^{3} h_{n,n+kN}^2\right)}_{c_n}\|\mathbf{a}_n\|_2^2 + \underbrace{\left(\sum_{\substack{i=1\\i\neq n}}^{N}\sum_{k=0}^{3} h_{n,n+kN}h_{i,n+kN}\mathbf{a}_i^T\right)}_{\mathbf{r}_n^{\mathrm{T}}}\mathbf{a}_n. \tag{A.2}$$

More explicitly, for a 4-neighborhood, the only terms in $\Phi(\mathbf{A})$ related to $\mathbf{a}_n$ can be expressed as

$$\begin{aligned}\phi(\mathbf{a}_n) =& \frac{1}{2}\underbrace{\big([n+W\leq n]+[n-W\geq 0]+[w\neq 0]+[w\neq 1]\big)}_{c_n}\|\mathbf{a}_n\|_2^2 \\ &- 2\underbrace{\big([n+W\leq n]\mathbf{a}_{n+W}+[n-W\geq 0]\mathbf{a}_{n-W}+[w\neq 0]\mathbf{a}_{n+1}+[w\neq 1]\mathbf{a}_{n-1}\big)^{\mathrm{T}}}_{\mathbf{r}_n^{\mathrm{T}}}\mathbf{a}_n\end{aligned} \tag{A.3}$$

where $[\mathscr{P}]$ corresponds to Iverson brackets of the logical expression $\mathscr{P}$, and

$$n = hW + w,\ h\in\{0,\dots,H\},\ w\in\{0,\dots,W-1\}. \tag{A.4}$$

A.1.2. Endmember penalization

Distance between the endmembers and reference signatures. The distance between the endmembers and the available reference signatures is

$$\Psi(\mathbf{M}) = \frac{1}{2}\|\mathbf{M}-\mathbf{M}_0\|_{\mathrm{F}}^2 = \frac{1}{2}\sum_{\ell=1}^{L}\|\widetilde{\mathbf{m}}_\ell-\widetilde{\mathbf{m}}_{\ell,0}\|_2^2\,. \tag{A.5}$$

As a consequence, the penalty for the ℓth band is

$$\psi(\widetilde{\mathbf{m}}_\ell) = \frac{1}{2}\|\widetilde{\mathbf{m}}_\ell-\widetilde{\mathbf{m}}_{\ell,0}\|_2^2\,. \tag{A.6}$$

Mutual distance between the endmembers. The distance between the different endmembers can be expressed as follows

$$\Psi(\mathbf{M}) = \frac{1}{2}\sum_{i=1}^{R}\left(\sum_{\substack{j=1\\j\neq i}}^{R}\|\mathbf{m}_i-\mathbf{m}_j\|_2^2\right) = \frac{1}{2}\sum_{r=1}^{R}\|\mathbf{M}\mathbf{G}_r\|_{\mathrm{F}}^2 = \frac{1}{2}\|\mathbf{M}\mathbf{G}\|_{\mathrm{F}}^2 \tag{A.7}$$

with

$$\mathbf{G} = \Big[\mathbf{G}_1, \ldots, \mathbf{G}_R\Big] \in \mathbb{R}^{R \times R^2}$$

and for $r = 1, \ldots, R$

$$\mathbf{G}_r = -\mathbf{I}_R + \mathbf{c}_r \mathbf{1}_R^{\mathrm{T}}$$

where $\mathbf{c}_r$ denotes the rth vector in the canonical basis of $\mathbb{R}^R$. Hence

$$\psi(\widetilde{\mathbf{m}}_\ell) = \frac{1}{2} \left\| \widetilde{\mathbf{m}}_\ell \mathbf{G} \right\|_2^2. \tag{A.8}$$

Volume penalization function. The volume penalization is expressed in terms of the projected endmembers $\mathbf{E}$, hence the need to find a condition equivalent to the positivity of both $\mathbf{M}$ and $\mathbf{M} + \mathbf{dM}_n$ (see [Dob+09]). We will first analyze the expression of the volume penalization with respect to $\widetilde{\mathbf{e}}_k$, and then give a condition on $\mathbf{E}$ ensuring the positivity of $\mathbf{M}$ (respectively $\mathbf{M} + \mathbf{dM}_n$ when endmember variability is considered). We first observe that the determinant of a matrix $\mathbf{X} \in \mathbb{R}^{R \times R}$ can be developed along its ith row, yielding

$$\det(\mathbf{X}) = \sum_j (-1)^{i+j} x_{ij} \det(\mathbf{X}_{\backslash i, \backslash j}) = \widetilde{\mathbf{x}}_i \mathbf{f}_i, \text{ with } \mathbf{f}_i = \left[(-1)^{i+j} \det(\mathbf{X}_{\backslash i, \backslash j})\right]_{j=1}^{R} \in \mathbb{R}^R.$$

Consequently, for $i = 1, \ldots, K$

$$\det \begin{pmatrix} \mathbf{E} \\ \mathbf{1}_R^T \end{pmatrix} = \widetilde{\mathbf{e}}_i \mathbf{f}_i.$$

Using the previous developments

$$\psi(\widetilde{\mathbf{e}}_i) = \frac{1}{2(R-1)!^2} (\widetilde{\mathbf{e}}_i \mathbf{f}_i)^2. \tag{A.9}$$

Positivity constraints and volume penalization. Using the fact that

$$\mathbf{M} = \mathbf{U}\mathbf{E} + \bar{\mathbf{y}} \mathbf{1}_R, \quad \text{with} \quad \mathbf{U}^{\mathrm{T}} \mathbf{U} = \mathbf{I}_K$$

one has

$$m_{\ell r} = \sum_j u_{\ell j} t_{jr} + \bar{y}_\ell. \tag{A.10}$$

The positivity constraints (1.3) on $m_{\ell r}$ can then be expressed as

$$e_{kr} \geq -\frac{\bar{y}_\ell + \sum_{j \neq k} u_{\ell j} e_{jr} + b_{\ell,r}}{u_{\ell k}}, \quad b_{\ell,r} = \min\big\{0, \min_n \{dm_{\ell,r,n}\}\big\}.$$

Introducing the two sets of integers

$$\mathcal{U}_k^+ = \{\ell | u_{\ell k} > 0\}, \quad \mathcal{U}_k^- = \{\ell | u_{\ell k} < 0\}$$

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

$$e_{kr}^- = \max_{\ell \in \mathcal{U}_k^+} \left(-\frac{\bar{y}_\ell + \sum_{j \neq k} u_{\ell j} e_{jr} + b_{\ell,r}}{u_{\ell k}} \right) \tag{A.11}$$

$$e_{kr}^+ = \min_{\ell \in \mathcal{U}_k^-} \left(-\frac{\bar{y}_\ell + \sum_{j \neq k} u_{\ell j} e_{jr} + b_{\ell,r}}{u_{\ell k}} \right). \tag{A.12}$$

A.1.3. Variability penalization

The variability energy penalty introduced in (1.12) is given by

$$\Upsilon \left(\mathbf{dM} \right) = \frac{1}{2} \left\| \mathbf{dM} \right\|_{\mathrm{F}}^2 \Rightarrow \upsilon \left(\mathbf{dM}_n \right) = \frac{1}{2} \left\| \mathbf{dM}_n \right\|_{\mathrm{F}}^2. \tag{A.13}$$

A.2. ADMM: general principle

The ADMM is a technique combining the benefits of augmented Lagrangian and dual decomposition methods to solve constrained optimization problems [Boy+10]. 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 [Boy+10]) 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

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

The scaled augmented Lagrangian associated with this problem can be written

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

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

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

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 [Boy+10, p. 19]): the procedure is iterated until

$$\| \mathbf{r}^q \|_2 \leq \varepsilon^{\text{pri}} \quad \text{and} \quad \| \mathbf{s}^q \|_2 \leq \varepsilon^{\text{dual}} \tag{A.15}$$

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

$$\mathbf{r}^{q+1} = \mathbf{A}\mathbf{x}^{q+1} + \mathbf{B}\mathbf{z}^{q+1} - \mathbf{c} \tag{A.16}$$

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

and

$$\varepsilon^{\text{pri}} = \sqrt{p}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\Big\{\|\mathbf{A}\mathbf{x}^{q}\|_2, \|\mathbf{B}\mathbf{z}^{q}\|_2, \|\mathbf{c}\|_2\Big\} \tag{A.18}$$

$$\varepsilon^{\text{dual}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\rho\|\mathbf{A}^{\mathrm{T}}\mathbf{u}^{q}\|_2. \tag{A.19}$$

Finally, the parameter ρ can be adjusted using the rule described in [Boy+10, p. 20]

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

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

A.3. Solutions to the ADMM optimization sub-problems

In the following paragraphs, the index k denotes the current iteration of the BCD algorithm (outer loop), whereas q refers to the iteration index of the ADMM algorithm considered (inner loops).

A.3.1. Abundance estimation

Using (A.3), the scaled augmented Lagrangian (1.14) becomes

$$\begin{aligned}\mathcal{L}_{\mu_n^{(\mathbf{A})}}\left(\mathbf{a}_n, \mathbf{w}_n^{(\mathbf{A})}, \boldsymbol{\lambda}_n^{(\mathbf{A})}, \lambda_n\right) = \frac{1}{2}\|\mathbf{y}_n - (\mathbf{M}^k + \mathbf{dM}_n^k)\mathbf{a}_n\|_2^2 + \frac{\mu_n^{(\mathbf{A})}}{2}\|\mathbf{a}_n - \mathbf{w}_n^{(\mathbf{A})} + \boldsymbol{\lambda}_n^{(\mathbf{A})}\|_2^2 \\ + \lambda_n(\mathbf{1}_R^{\mathrm{T}}\mathbf{a}_n - 1) + \alpha\phi(\mathbf{a}_n) + \iota_{\{\cdot\succeq\mathbf{0}_R\}}\left(\mathbf{w}_n^{(\mathbf{A})}\right)\end{aligned}$$

Thus, introducing

$$\mathbf{B}_n^q = \left[(\mathbf{M}_n^k)^{\mathrm{T}}\mathbf{M}_n^k + (\alpha c_n^k + \mu_n^{(\mathbf{A})})\mathbf{I}_R\right]^{-1}, \quad \mathbf{M}_n^k = \mathbf{M}^k + \mathbf{dM}_n^k \tag{A.21}$$

$$\mathbf{u}_n^q = \mathbf{M}_n^k\mathbf{y}_n + \alpha\mathbf{r}_n^k + \mu_n^{(\mathbf{A})}(\mathbf{w}_n^{(\mathbf{A})(q)} - \boldsymbol{\lambda}_n^{(\mathbf{A})(q)}) \tag{A.22}$$

for $n = 1, \ldots, N$, and using the fact that

$$\lambda_n^q = \frac{\mathbf{1}_R^{\mathrm{T}}(\mathbf{B}_n^q)^{-1}\mathbf{u}_n^q - 1}{\mathbf{1}_R^{\mathrm{T}}(\mathbf{B}_n^q)^{-1}\mathbf{1}_R} \tag{A.23}$$

leads to

$$\mathbf{a}_n^{(k,q+1)} = (\mathbf{B}_n^q)^{-1}\left(\mathbf{u}_n^q - \frac{\mathbf{1}_R^{\mathrm{T}}(\mathbf{B}_n^q)^{-1}\mathbf{u}_n^q - 1}{\mathbf{1}_R^{\mathrm{T}}(\mathbf{B}_n^q)^{-1}\mathbf{1}_R}\mathbf{1}_R\right). \tag{A.24}$$

The update of the splitting variable is finally given by

$$\mathbf{w}_n^{(\mathbf{A})(q+1)} = \max\left(\mathbf{a}_n^{(k,q+1)} + \mathbf{w}_n^{(\mathbf{A})(q)}, \mathbf{0}_R\right) \tag{A.25}$$

where the $\max$ must be understood as a term-wise operator. In the absence of any penalization, the solution is obtained by taking $\alpha = 0$ in the previous equations.

A.3.2. Endmember estimation

Distance between the endmembers and reference spectral signatures. Using (A.6), the scaled augmented Lagrangian (1.16) is

$$\begin{aligned}\mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}, \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right) = \frac{1}{2}\left\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A}^{k+1} - \widetilde{\boldsymbol{\delta}}_\ell^k\right\|_2^2 + \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})} + \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right\|_2^2 \\ + \frac{\beta}{2}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{m}}_{\ell,0}\right\|_2^2 + \iota_{\{\cdot\succeq\widetilde{\mathbf{c}}_\ell\}}\left(\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}\right)\end{aligned}$$

Thus

$$\widetilde{\mathbf{m}}_\ell^{(k,q+1)} = \left[\left(\widetilde{\mathbf{y}}_\ell - \widetilde{\boldsymbol{\delta}}_\ell^k\right)(\mathbf{A}^{k+1})^T + \beta\widetilde{\mathbf{m}}_{\ell,0} + \mu_\ell^{(\mathbf{M})}\left(\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})(q)} - \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})(q)}\right)\right]\left[\mathbf{A}^{k+1}(\mathbf{A}^{k+1})^{\mathrm{T}} + (\mu_\ell^{(\mathbf{M})} + \beta)\mathbf{I}_R\right]^{-1} \tag{A.26}$$

and

$$\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})(q+1)} = \max\left(\widetilde{\mathbf{m}}_\ell^{(k,q+1)} + \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})(q)}, \widetilde{\mathbf{c}}_\ell\right). \tag{A.27}$$

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

Mutual distance between the endmembers. Using (A.8), the scaled augmented Lagrangian (1.16) is

$$\begin{aligned}\mathcal{L}_{\mu_\ell^{(\mathbf{M})}}\left(\widetilde{\mathbf{m}}_\ell, \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}, \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right) = \frac{1}{2}\left\|\widetilde{\mathbf{y}}_\ell - \widetilde{\mathbf{m}}_\ell\mathbf{A}^{k+1} - \widetilde{\boldsymbol{\delta}}_\ell^k\right\|_2^2 + \frac{\mu_\ell^{(\mathbf{M})}}{2}\left\|\widetilde{\mathbf{m}}_\ell - \widetilde{\mathbf{w}}_\ell^{(\mathbf{M})} + \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})}\right\|_2^2 \\ + \frac{\beta}{2}\left\|\widetilde{\mathbf{m}}_\ell\mathbf{G}\right\|_2^2 + \iota_{\{\cdot\succeq\widetilde{\mathbf{c}}_\ell\}}\left(\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})}\right).\end{aligned}$$

Thus

$$\widetilde{\mathbf{m}}_\ell^{(k,q+1)} = \left[\left(\widetilde{\mathbf{y}}_\ell - \widetilde{\boldsymbol{\delta}}_\ell^k\right)(\mathbf{A}^{k+1})^{\mathrm{T}} = +\mu_\ell^{(\mathbf{M})}\left(\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})(q)} - \widetilde{\boldsymbol{\lambda}}_\ell^{(\mathbf{M})(q)}\right)\right]\left[\mathbf{A}^{k+1}(\mathbf{A}^{k+1})^{\mathrm{T}} + \beta\mathbf{G}\mathbf{G}^{\mathrm{T}} + \mu_\ell^{(\mathbf{M})}\mathbf{I}_R\right]^{-1} \tag{A.28}$$

with $\widetilde{\mathbf{w}}_\ell^{(\mathbf{M})(q+1)}$ given by (A.27).

Volume penalization. Since the penalty is expressed with respect to the projected endmembers $\mathbf{E}$, the optimization sub-problems related to the endmembers have to be re-written accordingly. We first note that

$$\begin{aligned}
\|\mathbf{Y} - \mathbf{M}\mathbf{A} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2 &= \|\mathbf{Y} - (\mathbf{U}\mathbf{E} - \bar{\mathbf{y}}\mathbf{1}_R)\mathbf{A} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2 \\
&= \|\mathbf{U}\mathbf{E}\mathbf{A}\|_{\mathrm{F}}^2 - 2\langle \mathbf{U}\mathbf{E}\mathbf{A}, \mathbf{Y} + \bar{\mathbf{y}}\mathbf{1}_R\mathbf{A} - \boldsymbol{\Delta}\rangle + \|\mathbf{Y} + \bar{\mathbf{y}}\mathbf{1}_R\mathbf{A} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2 \\
&= \|\mathbf{E}\mathbf{A}\|_{\mathrm{F}}^2 - 2\langle \mathbf{E}, \underbrace{\mathbf{U}^{\mathrm{T}}\big(\mathbf{Y} + \bar{\mathbf{y}}\mathbf{1}_R\mathbf{A} - \boldsymbol{\Delta}\big)\mathbf{A}^{\mathrm{T}}}_{\mathbf{S}}\rangle + \|\mathbf{Y} + \bar{\mathbf{y}}\mathbf{1}_R\mathbf{A} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2, \quad (\text{since } \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_K).
\end{aligned}$$

Since $\langle \mathbf{S}, \mathbf{E}\rangle = \sum_{i=1}^{K} \langle \widetilde{\mathbf{s}}_i, \widetilde{\mathbf{e}}_i\rangle$ the resulting sub-problems for $i = 1, \dots, K$ are given by

$$\widetilde{\mathbf{e}}_i^* = \underset{\widetilde{\mathbf{e}}_i}{\arg\min} \left\{ \frac{1}{2} \left\|\widetilde{\mathbf{e}}_i \mathbf{A}^{k+1}\right\|_2^2 - \langle \widetilde{\mathbf{s}}_i^k, \widetilde{\mathbf{e}}_i\rangle + \frac{\beta}{2(R-1)!^2} (\widetilde{\mathbf{e}}_i \mathbf{f}_i)^2, \text{ s.t. } e_{i,r} \in [e_{i,r}^-, e_{i,r}^+] \text{ for } r \in \{1, \dots, R\} \right\}$$

where $\widetilde{\mathbf{e}}_i$ denotes the ith row of $\mathbf{E}$. Introducing the splitting variables $\widetilde{\mathbf{w}}_i^{(\mathbf{E})}$ such that $\widetilde{\mathbf{e}}_i = \widetilde{\mathbf{w}}_i^{(\mathbf{E})}$ for $i = 1, \dots, K$, the scaled augmented Lagrangian is

$$\begin{aligned}
\mathcal{L}_{\mu_k^{(\mathbf{T})}}\left(\widetilde{\mathbf{e}}_i, \widetilde{\mathbf{w}}_i^{(\mathbf{E})}, \widetilde{\boldsymbol{\lambda}}_i^{(\mathbf{E})}\right) =& \frac{1}{2} \left\|\widetilde{\mathbf{e}}_i \mathbf{A}^{k+1}\right\|_2^2 - \langle \widetilde{\mathbf{s}}_i^k, \widetilde{\mathbf{e}}_i\rangle + \frac{\beta}{2(R-1)!^2} (\widetilde{\mathbf{e}}_i \mathbf{f}_i)^2 + \frac{\mu_k^{(\mathbf{T})}}{2} \left\|\widetilde{\mathbf{e}}_i - \widetilde{\mathbf{w}}_i^{(\mathbf{E})} + \widetilde{\boldsymbol{\lambda}}_i^{(\mathbf{E})}\right\|_2^2 \\
&+ \iota_{\{\cdot \succeq \widetilde{\mathbf{e}}_i^-\}}\big(\widetilde{\mathbf{w}}_i^{(\mathbf{E})}\big) + \iota_{\{\cdot \preceq \widetilde{\mathbf{e}}_i^+\}}\big(\widetilde{\mathbf{w}}_i^{(\mathbf{E})}\big).
\end{aligned} \tag{A.29}$$

The update of the primal variable $\widetilde{\mathbf{e}}_i$ is consequently given by

$$\widetilde{\mathbf{e}}_i^{(k,q+1)} = \left[\mu_k^{(\mathbf{T})} (\widetilde{\mathbf{w}}_i^{(\mathbf{E})(q)} - \widetilde{\boldsymbol{\lambda}}_i^{(\mathbf{E})(q)}) - \widetilde{\mathbf{s}}_i^k\right] \left[\mathbf{A}^{k+1} (\mathbf{A}^{k+1})^{\mathrm{T}} + \frac{\beta}{(R-1)!^2} \mathbf{f}_i \mathbf{f}_i^T + \mu_k^{(\mathbf{T})} \mathbf{I}_R\right]^{-1}. \tag{A.30}$$

Finally, the splitting variable is given for $r = 1, \dots, R$ by

$$\widetilde{\mathbf{w}}_i^{(\mathbf{E})(q+1)} = \min\left\{\max\left\{\widetilde{\mathbf{e}}_i^{(k,q+1)} + \widetilde{\boldsymbol{\lambda}}_i^{(\mathbf{E})(q)}, \widetilde{\mathbf{e}}_i^-\right\}, \widetilde{\mathbf{e}}_i^+\right\}. \tag{A.31}$$

A.3.3. Variability estimation

Using (A.13), the scaled augmented Lagrangian (1.18) is

$$\begin{aligned}
\mathcal{L}_{\mu_n^{(\mathbf{dM})}}\big(\mathbf{dM}_n, \mathbf{W}_n^{(\mathbf{dM})}, \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\big) = \frac{1}{2} \|\mathbf{y}_n - (\mathbf{M}^k + \mathbf{dM}_n)\mathbf{a}_n^k\|_2^2 + \gamma\upsilon(\mathbf{dM}_n) + \iota_{\{\cdot \succeq -\mathbf{M}^k\}}\big(\mathbf{W}_n^{(\mathbf{dM})}\big) \\
+ \frac{\mu_n^{(\mathbf{dM})}}{2} \left\|\mathbf{dM}_n - \mathbf{W}_n^{(\mathbf{dM})} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})}\right\|_{\mathrm{F}}^2.
\end{aligned}$$

Hence

$$\mathbf{dM}_n^{(k-1,q+1)} = \left[(\mathbf{y_n} - \mathbf{M}^k \mathbf{a}_n^k)(\mathbf{a}_n^k)^{\mathrm{T}} + \mu_n^{(\mathbf{dM})} \Big(\mathbf{W}_n^{(\mathbf{dM})(q)} - \mathbf{M}^k - \boldsymbol{\Lambda}_n^{(\mathbf{dM})(q)}\Big)\right] \left[\mathbf{a}_n^k (\mathbf{a}_n^k)^T + (\mu_n^{(\mathbf{dM})} + \gamma)\mathbf{I}_R\right]^{-1} \tag{A.32}$$

and

$$\mathbf{W}_n^{(\mathbf{dM})(q)} = \max\Big(\mathbf{dM}_n^{(k-1,q+1)} + \boldsymbol{\Lambda}_n^{(\mathbf{dM})(q)}, -\mathbf{M}^k\Big).$$

A.4. Proof of (1.24)

The following proposition (essentially adapted from [BST13]) can be easily adapted to the matrix functions considered in (1.24).

Proposition 4. *Let* $\mathbf{a} \in \mathbb{R}^n$ *and* $\varepsilon \in \mathbb{R}_+^*$. *Then*

$$\operatorname{prox}_{\iota_{\{\|\cdot\|_2 \leq \varepsilon\}} + \iota_{\{\cdot \succeq \mathbf{a}\}}}(\mathbf{y}) = \operatorname{prox}_{\iota_{\{\|\cdot\|_2 \leq \varepsilon\}}}\big(\mathbf{a} + [\mathbf{y} - \mathbf{a}]_+\big). \tag{A.33}$$

Proof. We first note that

$$\operatorname{prox}_{\iota_{\{\|\cdot\|_2 \leq \varepsilon\}} + \iota_{\{\cdot \succeq \mathbf{a}\}}}(\mathbf{y}) = \mathbf{a} + \operatorname{prox}_{\iota_{\{\|\cdot + \mathbf{a}\|_2 \leq \varepsilon\}} + \iota_{\{\cdot \succeq \mathbf{0}\}}}(\mathbf{y} - \mathbf{a}) \tag{A.34}$$

Denoting $\mathbf{u} = \mathbf{y} - \mathbf{a}$, the problem of interest then reduces to the computation of $\operatorname{prox}_{\iota_{\{\|\cdot + \mathbf{a}\|_2 \leq \varepsilon\}} + \iota_{\{\cdot \succeq \mathbf{0}\}}}(\mathbf{u})$. To this end, we adopt the notations and the proof introduced in [BST13, Proposition 4]. Given $\mathbf{x} \in \mathbb{R}^n$, let us consider

$$\|\mathbf{x}\|_+^2 = \sum_{i \in \mathcal{I}^+} x_i^2, \quad \|\mathbf{x}\|_-^2 = \sum_{i \in \mathcal{I}^-} x_i^2 \tag{A.35}$$

where

$$\mathcal{I}^+ = \big\{i \in \{1, \ldots, n\} : x_i \geq 0\big\}, \quad \mathcal{I}^- = \big\{i \in \{1, \ldots, n\} : x_i < 0\big\}. \tag{A.36}$$

Note that (see [BST13])

(i) $\|\mathbf{x}\|_2^2 = \|\mathbf{x}\|_+^2 + \|\mathbf{x}\|_-^2$;

(ii) $\|\mathbf{x} - \mathbf{u}\|_+^2 + \|\mathbf{x}\|_-^2 = \|\mathbf{x} - [\mathbf{u}]_+\|_2^2$;

(iii) $\|\mathbf{x}\|_-^2 = 0 \Leftrightarrow x_i = 0 \, \forall i \in \mathcal{I}^-$.

With the above notations, the preceding remarks lead to

$$\begin{aligned}
\bar{\mathbf{x}} &\in \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\mathbf{u}\|_2^2, \quad \mathbf{x}\succeq\mathbf{0} \right\} \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\big(\|\mathbf{x}-\mathbf{u}\|_+^2 + \|\mathbf{x}-\mathbf{u}\|_-^2\big), \quad \mathbf{x}\succeq\mathbf{0} \right\} \quad \text{(according to (i))} \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\Big(\|\mathbf{x}-\mathbf{u}\|_+^2 + \|\mathbf{x}\|_-^2 - 2\sum_{i\in\mathcal{I}^-} x_i u_i\Big), \quad \mathbf{x}\succeq\mathbf{0} \right\} \quad (A.37) \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\mathbf{u}\|_+^2, \quad x_i = 0 \; \forall i\in\mathcal{I}^-, \quad \mathbf{x}\succeq\mathbf{0} \right\} \quad (A.38) \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\mathbf{u}\|_+^2, \quad \|\mathbf{x}\|_-^2 = 0 \right\} \quad \text{(according to (iii))} \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\big(\|\mathbf{x}-\mathbf{u}\|_+^2 + \|\mathbf{x}\|_-^2\big) \right\} \\
&= \underset{\mathbf{x}}{\arg\min} \left\{ \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-[\mathbf{u}]_+\|_2^2 \right\} \\
&= \operatorname{prox}_{\iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}}\big([\mathbf{u}]_+\big) \quad \text{(according to (ii))} \quad (A.39)
\end{aligned}$$

where the equality between (A.37) and (A.38) results from the following considerations. We first note that any solution to the problem (A.38) is a solution to (A.37). The converse implication follows by a contradiction argument. Let $\bar{\mathbf{x}}$ be a solution to (A.37), but not to (A.38). Then,

$$\begin{cases} \forall\mathbf{x}\succeq\mathbf{0}: \quad \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\bar{\mathbf{x}}) + \frac{1}{2}\Big(\|\bar{\mathbf{x}}-\mathbf{u}\|_+^2 + \|\bar{\mathbf{x}}\|_-^2 + \sum_{i\in\mathcal{I}^-}\bar{x}_i u_i\Big) \le \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\Big(\|\mathbf{x}-\mathbf{u}\|_+^2 + \|\mathbf{x}\|_-^2 + \sum_{i\in\mathcal{I}^-} x_i u_i\Big) \\ \exists\widetilde{\mathbf{x}}\succeq\mathbf{0}: \; (\forall i\in\mathcal{I}^-, \; \widetilde{x}_i = 0), \quad \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\widetilde{\mathbf{x}}) + \frac{1}{2}\|\widetilde{\mathbf{x}}-\mathbf{u}\|_+^2 < \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\bar{\mathbf{x}}-\mathbf{u}\|_+^2. \end{cases} \tag{A.40}$$

In particular, (A.40) implies

$$\iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\underbrace{\Big(\|\bar{\mathbf{x}}-\mathbf{u}\|_+^2 + \|\bar{\mathbf{x}}\|_-^2 + \sum_{i\in\mathcal{I}^-}\bar{x}_i u_i\Big)}_{=\|\bar{\mathbf{x}}-\mathbf{u}\|_2^2} \le \iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}(\mathbf{x}) + \frac{1}{2}\|\bar{\mathbf{x}}-\mathbf{u}\|_+^2 < \|\bar{\mathbf{x}}+\mathbf{a}\|_2 + \frac{1}{2}\|\bar{\mathbf{x}}-\mathbf{u}\|_+^2$$

leading to the following contradiction

$$\|\bar{\mathbf{x}}-\mathbf{u}\|_2^2 < \|\bar{\mathbf{x}}-\mathbf{u}\|_+^2.$$

To conclude, we note that

$$\operatorname{prox}_{\iota_{\{\|\cdot+\mathbf{a}\|_2\le\varepsilon\}}}(\mathbf{x}) = \operatorname{prox}_{\iota_{\{\|\cdot\|_2\le\varepsilon\}}}(\mathbf{x}+\mathbf{a}) - \mathbf{a} \tag{A.41}$$

and combine (A.34), (A.39) and (A.41) to obtain

$$\operatorname{prox}_{\iota_{\{\|\cdot\|_2\le\varepsilon\}}+\iota_{\{\cdot\succeq\mathbf{a}\}}}(\mathbf{y}) = \operatorname{prox}_{\iota_{\{\|\cdot\|_2\le\varepsilon\}}}\big(\mathbf{a}+[\mathbf{y}-\mathbf{a}]_+\big).$$

□

Remark A.1. Note that [Yu13, Theorem 4] was not directly applicable in this context, since $\iota_{\{\cdot \succeq \mathbf{a}\}}$ is not positively homogeneous (up to a constant term).

APPENDIX B

Appendix to chapter 2

B.1. Details on the values chosen in Table 2.2

The initial values chosen for the parameters in Table 2.2 are based on the following considerations.

(a) The initial noise variance σ_t^2 reflects a typical order of magnitude of the SNR (HS images are typically known to have an SNR between 30 and 35 dB).

(b) The initial outlier variance s_t^2 has been taken an order of magnitude greater than σ_t^2 to ensure outlier contributions that can be captured by the algorithm.

(c) The initial variability variance $\psi_{\ell,r}^2$, which controls the temporal smoothness of the variability term, can be *a priori* chosen of the same order of magnitude as s_t^2.

(d) The granularity parameters β_t were initially selected to reflect the practitioner's prior knowledge on the smoothness of the outlier spatial support. A value between 1 and 2 (i.e., above the phase-transition temperature of the Ising MRF [Per+14]) is particularly appropriate for natural scenes, in which the observed materials exhibit a relatively smooth spatial distribution.

Similarly, the values of the fixed parameters given in Table 2.2 are selected as follows.

(a) ε_n^2, which controls the confidence given to the abundance smoothness prior, has been fixed by cross-validation (i.e., based on the estimation results obtained by multiple runs for different values);

(b) since no specific prior knowledge is available on the endmembers, the endmember variance ξ is chosen sufficiently large (typically equal to 1) to ensure the endmember prior is weakly informative;

(c) ν, which controls the energy of the variability captured by the algorithm, has been set by cross validation;

(d) given the absence of specific prior knowledge on the outlier variances s_t^2, the variability variances $\psi_{\ell,r}^2$ and the noise variances σ_t^2, the hyperparameters $a_{\mathbf{s}}, a_{\Psi}, a_{\sigma}, b_{\mathbf{s}}, b_{\Psi}, b_{\sigma}$ are set to a small value (typically 10^{-3}) to ensure the chosen conjugate inverse-gamma priors are uninformative [AMH15];

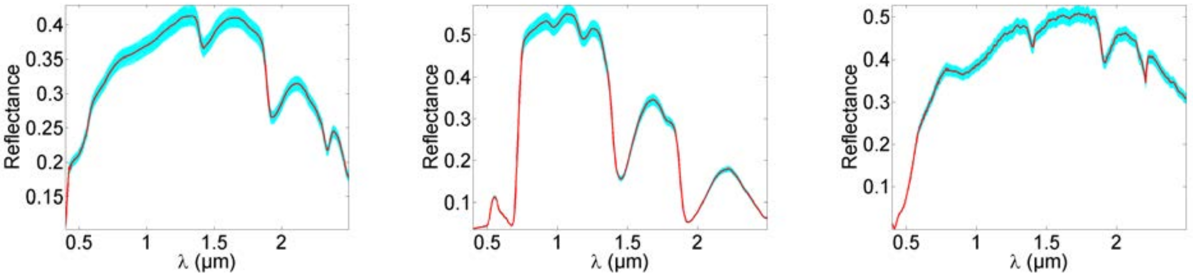
(e) the number of Monte-Carlo and burn-in iterations, respectively denoted by N_{MC} and N_{bi}, are set according to a classical convergence diagnosis, namely the potential scale reduction factor (PSRF) [GR92].

B.2. Results for the the second scenario (datasets $\sharp 4$ and $\sharp 5$)

The synthetic datasets $\sharp 4$ and $\sharp 5$ are affected by moderate spatial variability (i.e., variability within each image of the sequence, generated in the bottom left-hand corner of each image) and outliers (see Fig. B.2). An example of the spatial variability considered in these datasets is reported in Fig. B.1. The data have been further corrupted by an additive white Gaussian noise to ensure a resulting SNR between 30 and 35 dB. In the following paragraphs, only the results obtained for the mixtures composed of $R = 3$ endmembers are commented (Table B.1 and Figs. B.3 to B.6). Similar comments can be made for the case $R = 9$. Illustrations related to this latter case are omitted for the sake of brevity.

- **Endmember estimation:** the proposed method shows an interesting robustness with respect to spatially sparse outliers and moderate spatial variability, in the sense that the estimated signatures (Figs. B.6(p) to B.6(r)) are very close to the corresponding ground truth (Fig. B.1). In comparison, the shape of the endmembers recovered by VCA, SISAL and RLMM and the variability extracted by OU are significantly affected by outliers, as exemplified in Figs. B.6(e), B.6(h), B.6(k) and B.6(n) respectively. These qualitative results are confirmed by the quantitative performance measures reported in Table B.1.
- **Abundance estimation:** the abundance maps estimated by FCLS, RLMM and SISAL reflect the high sensitivity of VCA (used to initialize SISAL and RLMM) to the presence of outliers (see the figures delineated in red in Figs. B.3 to B.5). On the contrary, the abundances recovered by OU and the proposed approach are much closer to the ground truth. These observations are confirmed by the abundance estimation performance reported in Table B.1.
- **Overall performance:** the performance measures reported in Table B.1 are globally favorable to the proposed approach. It is important to mention that the price to pay with the good performance of the proposed method is its computational complexity, which is common with MCMC methods.

We finally observed that spatial variability has an influence on the outlier maps estimated by the algorithm (Fig. B.2): as naturally expected, more pixels tend to be detected as corrupted by outliers. Indeed, for a given time instant t, spatial variability induces local deviations from the linear model $(\mathbf{M} + \mathbf{dM}_t)\mathbf{A}_t$, which are then captured as outliers or noise by the proposed model. Note however that this phenomenon does not significantly affect the recovered abundances and endmembers, as illustrated in Figs. B.3-B.5 and B.6.



(a) True endmember 1 (b) True endmember 2 (c) True endmember 3

Figure B.1.: True endmembers (in red) and spatial variability (in cyan) used for the synthetic dataset with $R = 3$.

Table B.1.: Simulation results on synthetic data affected by moderate spatial variability (GMSE($\mathbf{A}$)$\times 10^{-2}$, GMSE($\mathbf{dM}$)$\times 10^{-4}$, RE $\times 10^{-4}$).

		aSAM($\mathbf{M}$) (°)	GMSE($\mathbf{A}$)	GMSE($\mathbf{dM}$)	RE	time (s)
$R = 3$ (♯4)	VCA/FCLS	4.83	2.10	/	2.39	**1**
	SISAL/FCLS	4.71	1.55	/	0.82	2
	rLMM	3.92	1.68	/	**0.22**	655
	OU	1.46	0.35	3.22	0.95	113
	Proposed (MCMC)	**1.31**	**0.09**	**0.33**	0.65	2425
$R = 9$ (♯5)	VCA/FCLS	2.93	3.12 $\times 10^{-1}$	/	20.1	**5**
	SISAL/FCLS	5.89	4.67 $\times 10^{-1}$	/	0.87	6
	rLMM	2.74	2.83 $\times 10^{-1}$	/	**0.38**	1498
	OU	2.48	**1.63** $\times 10^{-1}$	**2.81**	1.20	379
	Proposed (MCMC)	**1.60**	2.00 $\times 10^{-1}$	5.65	0.84	8926

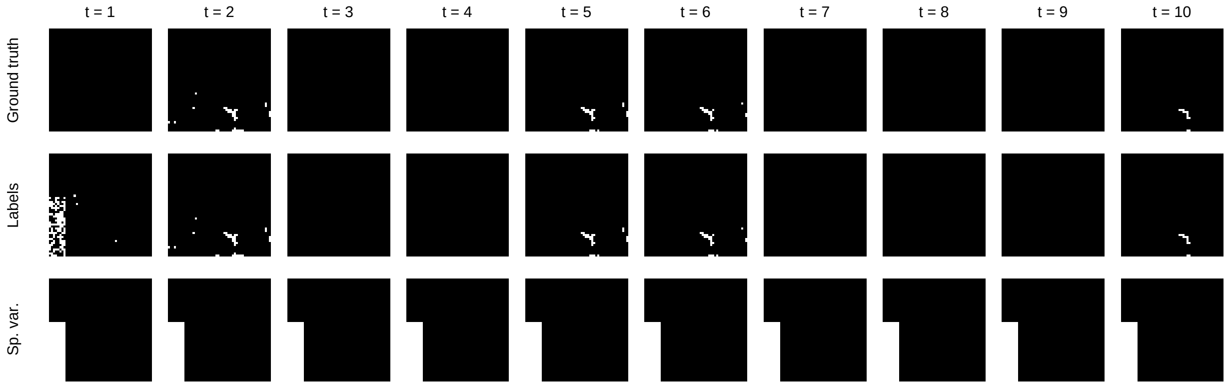


Figure B.2.: mMAP estimates of the label maps recovered by the proposed approach, displayed at each time instant (the different rows correspond to the true label map where outliers appear in white, the estimated map and the areas where spatial variability is present). The influence of spatial variability can be seen on the outlier map estimated for the first image.

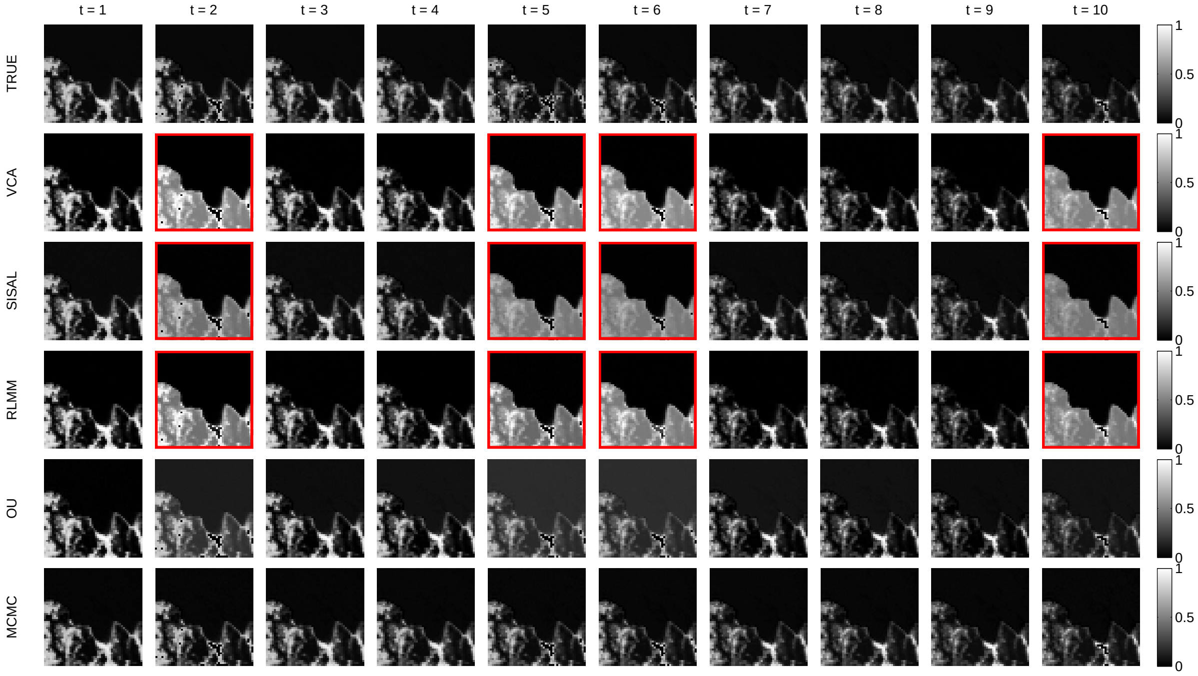


Figure B.3.: Abundance map of the endmember 1 recovered by the different methods (in each row) at each time instant (given in column) for the experiment with $R = 3$. The images delineated in red show that several methods are highly sensitive to the presence of outliers.

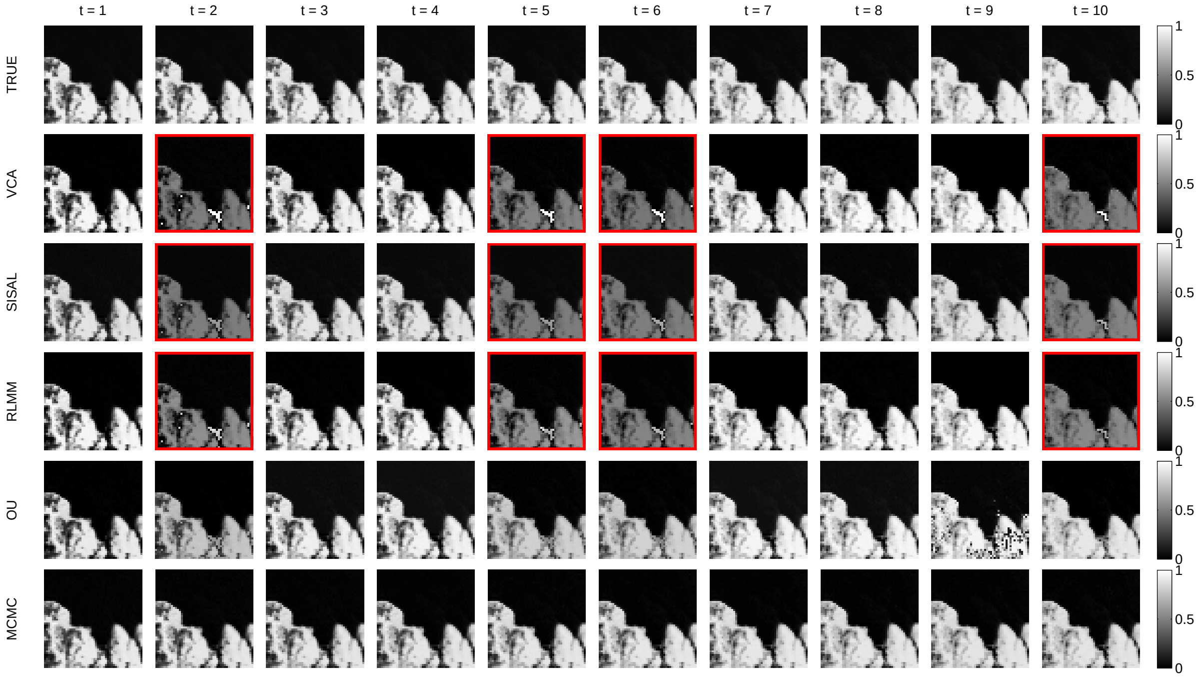


Figure B.4.: Abundance map of the endmember 2 recovered by the different methods (in each row) at each time instant (given in column) for the experiment with $R = 3$. The images delineated in red show that several methods are highly sensitive to the presence of outliers.

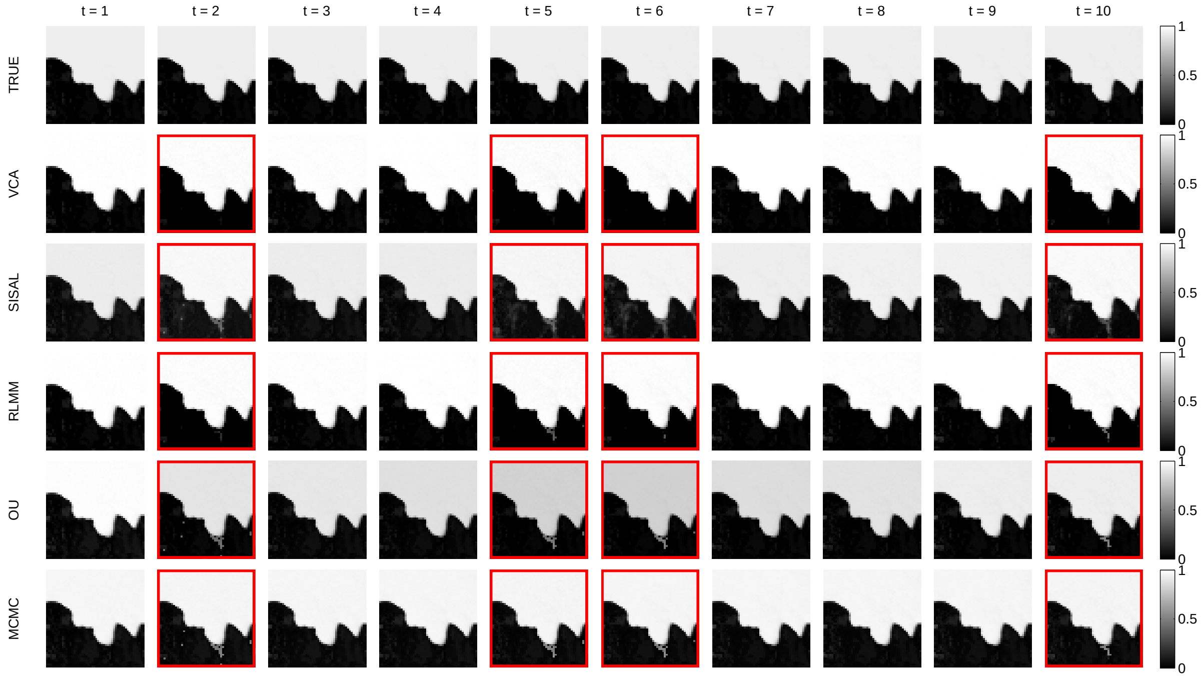
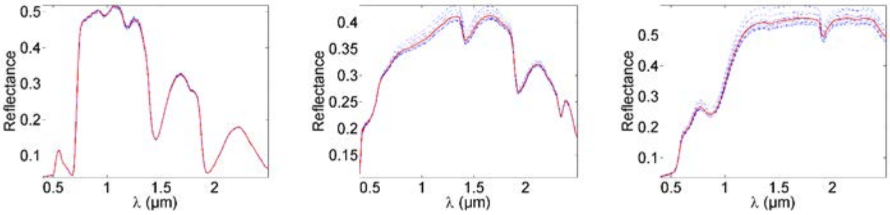
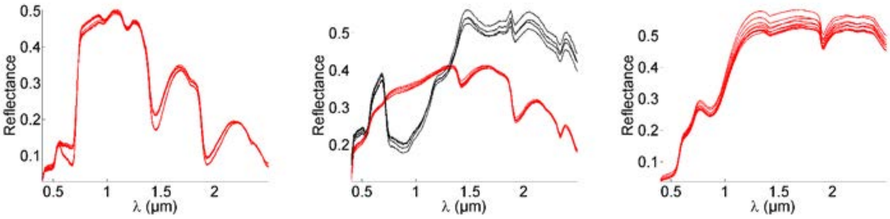


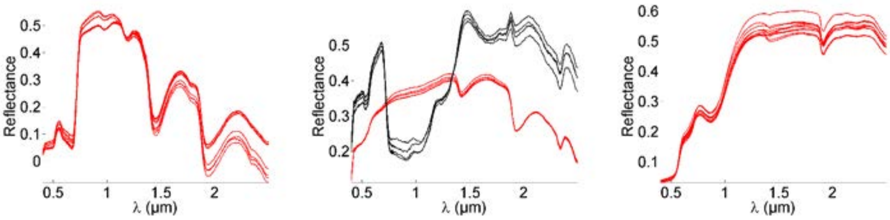
Figure B.5.: Abundance map of the endmember 3 recovered by the different methods (in each row) at each time instant (given in column) for the experiment with $R = 3$. The images delineated in red show that several methods are highly sensitive to the presence of outliers.



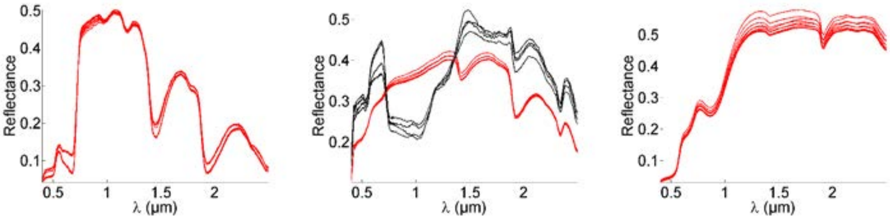
(a) True endmembers (b) True endmembers (c) True endmembers



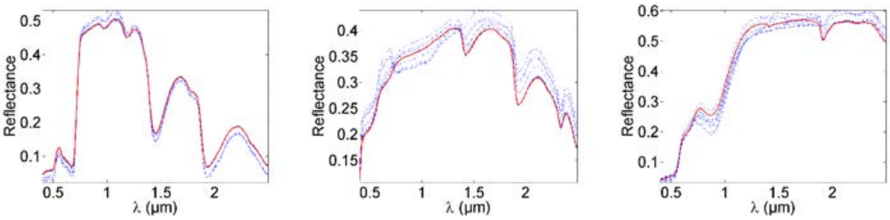
(d) VCA (e) VCA (f) VCA



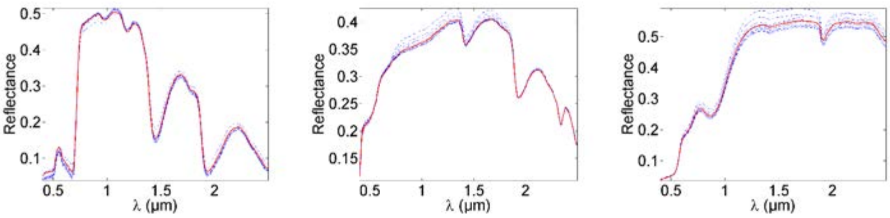
(g) SISAL (h) SISAL (i) SISAL



(j) RLMM (k) RLMM (l) RLMM



(m) OU (n) OU (o) OU



(p) MCMC (q) MCMC (r) MCMC

Figure B.6.: Endmembers ($\mathbf{m}_r$, red lines) and their variants affected by variability ($\mathbf{m}_r + \mathbf{dm}_{r,t}$, blue dotted lines) recovered by the different methods from the synthetic dataset corrupted by spatial variability ($R = 3$). Signatures corresponding to different time instants are represented on a single figure to better appreciate the variability recovered from the data. The spectra represented in black correspond to signatures significantly affected by outliers.

APPENDIX C

Appendix to chapter 3

C.1. Projections involved in the parameter updates

The projections involved in the PALM algorithm [BST13] described in Algo. 6 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 [BD86; HAS13]. Besides, the projection on a Frobenius ball is given by [PB14]
$$\mathcal{P}_{\mathcal{B}_{\mathrm{F}}(\mathbf{X},r)}(\mathbf{Y}) = \mathbf{X} + \min\left(1, \frac{r}{\|\mathbf{Y}-\mathbf{X}\|_{\mathrm{F}}}\right)(\mathbf{Y}-\mathbf{X}); \tag{C.1}$$
- projecting $\mathbf{M}$ onto $\mathbb{R}_+^{L\times R}$ is explicitly given by
$$\mathcal{P}_+(\mathbf{M}) = \max(\mathbf{0}_{L,R}, \mathbf{M}) \tag{C.2}$$
where the $\max$ is taken term-wise.

C.2. 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

$$\mathbf{H}_{(\mathbf{A},\mathbf{dM})}f = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \end{bmatrix} \tag{C.3}$$
$$\widetilde{\mathbf{M}} = (\mathbf{M}+\mathbf{dM}) \tag{C.4}$$
$$\mathbf{H}_1 = \mathbf{I}_N \otimes (\widetilde{\mathbf{M}}^{\mathrm{T}}\widetilde{\mathbf{M}}), \quad \mathbf{H}_4 = (\mathbf{A}\mathbf{A}^{\mathrm{T}}) \otimes \mathbf{I}_L \tag{C.5}$$
$$\mathbf{H}_3 = \mathbf{H}_2^{\mathrm{T}} = \left\{\mathbf{I}_R \otimes \left[-\mathbf{Y} + \widetilde{\mathbf{M}}\mathbf{A}\right]\right\}\mathbf{S}_{R,N} + [\mathbf{A} \otimes \widetilde{\mathbf{M}}]. \tag{C.6}$$

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 Schur complement of $\mathbf{H}_1$ can be computed more explicitly, leading to

$$\begin{aligned}\mathbf{H}_4 - \mathbf{H}_3\mathbf{H}_1^{-1}\mathbf{H}_2 = &\left[(\mathbf{A}\mathbf{A}^{\mathrm{T}}) \otimes (\mathbf{I}_L - \widetilde{\mathbf{M}}^{\dagger})\right] \\ &- \left[(\widetilde{\mathbf{M}}^{\mathrm{T}}\widetilde{\mathbf{M}})^{-1} \otimes (\mathbf{Y} - \widetilde{\mathbf{M}}\mathbf{A})(\mathbf{Y} - \widetilde{\mathbf{M}}\mathbf{A})^{\mathrm{T}}\right] \\ &+ 2\mathscr{S}\left(\mathbf{S}_{R,L}\left[(\mathbf{Y} - \widetilde{\mathbf{M}}\mathbf{A})\mathbf{A}^{\mathrm{T}} \otimes \widetilde{\mathbf{M}}^{\dagger}\right]\right)\end{aligned} \tag{C.7}$$

where $\mathscr{S}(\mathbf{X})$ denotes the symmetric part of the square matrix $\mathbf{X}$, $\widetilde{\mathbf{M}}^{\dagger}$ is the pseudo-inverse of $\widetilde{\mathbf{M}}$ and $\mathbf{S}_{R,L}$ is the perfect shuffle matrix defined by

$$\mathbf{S}_{R,L} = \sum_i \sum_j \mathbf{E}_{ij}(R,L)\mathbf{E}_{ij}(R,L)^{\mathrm{T}} \tag{C.8}$$

where $\mathbf{E}_{ij}(R,L)$ is the element (i,j) of the canonical basis of $\mathbb{R}^{R\times L}$. The invertibility of the Schur complement $\mathbf{S}$ can be ensured via an appropriate regularization term $\frac{\mu}{2}\left\|\mathbf{A}\right\|_{\mathrm{F}}^2$ 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 tothe 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 $\frac{\mu}{2}\left\|\mathbf{A}\right\|_{\mathrm{F}}^2$ 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.

C.3. Convergence proof

Largely adapted from [Mai+10], the following sketch of proof reduces to an adaptation of [Mai+10, Lemma 1, Proposition 1]. From this point, our problem exactly satisfies the assumptions required to apply the same arguments as in [Mai+10, Proposition 2, Proposition 3], leading to the announced convergence result.

Lemma 1 (Asymptotic variations of $\mathbf{M}_t$ [Mai+10])**.** *Under Assumptions 1 and 2, we have*

$$\left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_F = O\left(\frac{1}{t}\right) \textit{ almost surely } (a.s.). \tag{C.9}$$

Proof. According to Assumption 1, $\hat{g}_t$ is strictly convex with a Hessian lower-bounded by a scalar $\mu_{\mathrm{M}} > 0$. Consequently, $\hat{g}_t$ satisfies the second-order growth condition

$$\hat{g}_t(\mathbf{M}^{(t+1)}) - \hat{g}_t(\mathbf{M}^{(t)}) \geq \mu_M \left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_{\mathrm{F}}^2. \tag{C.10}$$

Besides, since $\mathbf{M} \in [0,1]^{L\times R}$, we have $\left\|\mathbf{M}\right\|_{\mathrm{F}} \leq \sqrt{LR}$. Hence $\hat{g}_t$ is Lipschitz continuous with constant $c_t = \frac{1}{t}\left(\left\|\mathbf{D}_t\right\|_{\mathrm{F}} + \sqrt{LR}\left\|\mathbf{C}_t\right\|_{\mathrm{F}}\right)$

βc_Ψ. Indeed, given two matrices $\mathbf{M}_1, \mathbf{M}_2 \in [0,1]^{L\times R}$, we have

$$\begin{aligned}
|\hat{g}_t(\mathbf{M}_1) - \hat{g}_t(\mathbf{M}_2)| &\leq \beta \big|\Psi(\mathbf{M}_1) - \Psi(\mathbf{M}_2)\big| + \frac{1}{t}\left|\frac{1}{2}\langle \mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2, \mathbf{C}_t\rangle - \langle \mathbf{M}_1 - \mathbf{M}_2, \mathbf{D}_t\rangle\right| \\
&\leq \beta c_\Psi \left\|\mathbf{M}_1 - \mathbf{M}_2\right\|_{\mathrm{F}} + \frac{1}{t}\left\|\mathbf{M}_1 - \mathbf{M}_2\right\|_{\mathrm{F}} \left\|\mathbf{D}_t\right\|_{\mathrm{F}} \\
&\quad + \frac{1}{2t}\left\|\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2\right\|_{\mathrm{F}} \left\|\mathbf{C}_t\right\|_{\mathrm{F}}
\end{aligned} \tag{C.11}$$

where $\mathbf{C}_t$ and $\mathbf{D}_t$ are defined in (3.11). In addition

$$\begin{aligned}
\left\|\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 - \mathbf{M}_2^{\mathrm{T}}\mathbf{M}_2\right\|_{\mathrm{F}} &= \frac{1}{2}\|(\mathbf{M}_1+\mathbf{M}_2)^{\mathrm{T}}(\mathbf{M}_1-\mathbf{M}_2) + (\mathbf{M}_1-\mathbf{M}_2)^{\mathrm{T}}(\mathbf{M}_1+\mathbf{M}_2)\|_{\mathrm{F}} \\
&\leq 2\sqrt{LR}\left\|\mathbf{M}_1 - \mathbf{M}_2\right\|_{\mathrm{F}}
\end{aligned} \tag{C.12}$$

hence

$$|\hat{g}_t(\mathbf{M}_1) - \hat{g}_t(\mathbf{M}_2)| \leq c_t \left\|\mathbf{M}_1 - \mathbf{M}_2\right\|_{\mathrm{F}}. \tag{C.13}$$

Combining (C.10) and (C.13), we have

$$\left\|\mathbf{M}^{(t+1)} - \mathbf{M}^{(t)}\right\|_{\mathrm{F}} \leq \frac{c_t}{\mu_{\mathbf{M}}}. \tag{C.14}$$

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. □

In the rest of this appendix, the following compact notation will be used to denote the following Jacobian matrices of a function $f : \mathbb{R}^{m_1\times n_1} \times \mathbb{R}^{m_2\times n_2} \times \mathbb{R}^{m_3\times n_3} \to \mathbb{R}$:

$$\begin{aligned}
\mathbf{J}_f^{\mathbf{X}}(\mathbf{X},\mathbf{Y},\mathbf{Z}) &\overset{\text{def}}{=} \frac{\partial f}{\partial(\operatorname{vec}^{\mathrm{T}}\mathbf{X})}(\mathbf{X},\mathbf{Y},\mathbf{Z}) \in \mathbb{R}^{1\times m_1 n_1} \\
\mathbf{J}_f^{\mathbf{X},\mathbf{Y}}(\mathbf{X},\mathbf{Y},\mathbf{Z}) &\overset{\text{def}}{=} \frac{\partial f}{\partial(\operatorname{vec}^{\mathrm{T}}\mathbf{X})\partial(\operatorname{vec}^{\mathrm{T}}\mathbf{Y})}(\mathbf{X},\mathbf{Y},\mathbf{Z}) \in \mathbb{R}^{m_2 n_2\times m_1 n_1}.
\end{aligned}$$

Proposition 5 (Adapted from [Mai+10])**.** *We assume that the requirements in Assumption 1 to 3 are satisfied. Let* $(\mathbf{Y}_t, \tilde{\mathbf{M}})$ *be an element of* $\mathcal{Y} \times \mathcal{M}$*. Let us define*

$$\mathcal{Z}_t = \mathcal{A}_R \times \mathcal{D}_t \tag{C.15}$$

$$\mathcal{Q}(\mathbf{Y}_t, \tilde{\mathbf{M}}) = \Big\{(\mathbf{A}, \mathbf{dM}) \mid \mathbf{0} \in \nabla_{(\mathbf{A},\mathbf{dM})} f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}, \mathbf{dM}) + \big(\mathcal{Z}_t\big)^{\perp}_{(\mathbf{A},\mathbf{dM})}\Big\} \tag{C.16}$$

$$(\mathbf{A}_t^k, \mathbf{dM}_t^k) \in \mathcal{Q}(\mathbf{Y}_t, \tilde{\mathbf{M}}) \tag{C.17}$$

$$v(\mathbf{Y}_t, \tilde{\mathbf{M}}) = f\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k\big). \tag{C.18}$$

Then

1. *the function* $v(\mathbf{Y}_t, \cdot)$ *is continuously differentiable with respect to* $\mathbf{M}$ *on a neighborhood* V *of* $\tilde{\mathbf{M}}$*. Furthermore, if* $(\mathbf{A}_t^k, \mathbf{dM}_t^k) \in \operatorname{int}\mathcal{Z}_t$*, then* $\nabla_{\mathbf{M}} v(\mathbf{Y}_t, \tilde{\mathbf{M}}) = \nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k\big)$*;*

2. *g defined in (3.6) is continously differentiable and $\nabla_{\mathbf{M}} g(\tilde{\mathbf{M}}) = \mathbb{E}_{\mathbf{Y}_t}\big[\nabla_{\mathbf{M}} v(\mathbf{Y}_t, \tilde{\mathbf{M}})\big]$;*

3. *$\nabla_{\mathbf{M}} g$ is Lipschitz continuous.*

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

In practice, the PALM algorithm used in Algo. 5 yields a point $(\mathbf{A}_t^k, \mathbf{dM}_t^k)$ such that

$$\mathbf{0} \in \nabla_{(\mathbf{A},\mathbf{dM})} f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k) + \big(\mathcal{Z}_t\big)^{\perp}_{(\mathbf{A}_t^k, \mathbf{dM}_t^k)} \tag{C.19}$$

swhere $\big(\mathcal{Z}_t\big)^{\perp}_{(\mathbf{A}_t^k, \mathbf{dM}_t^k)}$ is the cone normal to $\mathcal{Z}_t$ at $(\mathbf{A}_t^k, \mathbf{dM}_t^k)$. In the following paragraphs, two cases are separately addressed.

First case. If $(\mathbf{A}_t^k, \mathbf{dM}_t^k) \in \operatorname{int} \mathcal{Z}_t$, we have $\big(\mathcal{Z}_t\big)^{\perp}_{(\mathbf{A}_t^k, \mathbf{dM}_t^k)} = \{\mathbf{0}\}$. Thus, $\nabla_{(\mathbf{A},\mathbf{dM})} f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k) = \mathbf{0}$. Using Assumption 3, the first statement then follows from the implicit function theorem [Lan99, Theorem 5.9 p.19]: there exists two open subsets $V \subset \mathcal{M}$, $W \subset \mathcal{Z}_t$ and a continuously differentiable function $\varphi : V \to W$ such that

(i) $\big(\tilde{\mathbf{M}}, (\mathbf{A}_t^k, \mathbf{dM}_t^k)\big) \in V \times W \subset \mathcal{M} \times \mathcal{Z}_t$;

(ii) for all $(\mathbf{M}, (\mathbf{A}, \mathbf{dM})) \in V \times W$, we have

$$[\nabla_{(\mathbf{A},\mathbf{dM})} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \mathbf{0}] \Rightarrow [(\mathbf{A}, \mathbf{dM}) = \varphi(\mathbf{M})]; \tag{C.20}$$

(iii)

$$\mathbf{J}_{\varphi}(\mathbf{Y}_t, \tilde{\mathbf{M}}) = -\underbrace{\mathbf{H}^{-1}_{(\mathbf{A},\mathbf{dM})} f\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[(L+N)R \times (L+N)R]} \underbrace{\mathbf{J}_f^{\mathbf{M},(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[(L+N)R \times LR]}. \tag{C.21}$$

Consequently, the function $v(\mathbf{Y}_t, \cdot)$ is $\mathcal{C}^1$, and

$$\underbrace{\mathbf{J}_v^{\mathbf{M}}(\mathbf{Y}_t, \tilde{\mathbf{M}})}_{[1 \times LR]} = \mathbf{J}_f^{\mathbf{M}}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big) + \underbrace{\mathbf{J}_f^{(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[1 \times (L+N)R]} \underbrace{\mathbf{J}_{\varphi}(\mathbf{Y}_t, \tilde{\mathbf{M}})}_{[(L+N)R \times LR]}. \tag{C.22}$$

The equation (C.22) finally leads to the conclusion that $v(\mathbf{Y}_t, \cdot)$ is $\mathcal{C}^1$. The second part of the first claim follows by observing that

$$\Big[\nabla_{(\mathbf{A},\mathbf{dM})} f(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \mathbf{0} \Leftrightarrow \mathbf{J}_f^{(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \mathbf{M}, \varphi(\mathbf{M})\big) = \mathbf{0}\Big] \Rightarrow \mathbf{J}_v^{\mathbf{M}}(\mathbf{Y}_t, \mathbf{M}) = \mathbf{J}_f^{\mathbf{M}}\big(\mathbf{Y}_t, \mathbf{M}, \varphi(\mathbf{M})\big).$$

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

To prove the third claim, it is sufficient to show that $\mathbf{J}_f^{\mathbf{M}}\big(\mathbf{Y}_t, \cdot, \varphi(\cdot)\big)$ (or equivalently $\nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \cdot, \varphi(\cdot)\big)$) is Lipschitz continuous on V. To this end, we first observe that $\mathbf{J}_{\varphi}(\mathbf{Y}_t, \mathbf{M})$ is upper-bounded by a constant independent from $\mathbf{Y}_t$, which implies that φ is Lipschitz continuous. Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be two elements in V, and define $(\mathbf{A}_i, \mathbf{dM}_i) = \varphi(\mathbf{M}_i)$

for $i = 1, 2$. Then

$$\begin{aligned}
&\left\|\nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big) - \nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \mathbf{M}_2, \varphi(\mathbf{M}_2)\big)\right\|_{\mathrm{F}} \\
&= \left\|(\mathbf{M}_1 + \mathbf{dM}_1)\mathbf{A}_1\mathbf{A}_1^{\mathrm{T}} - (\mathbf{M}_2 + \mathbf{dM}_2)\mathbf{A}_2\mathbf{A}_2^{\mathrm{T}} - \mathbf{Y}_t(\mathbf{A}_1 - \mathbf{A}_2)^{\mathrm{T}}\right\|_{\mathrm{F}} \\
&= \Big\|\big[(\mathbf{M}_1 - \mathbf{M}_2) + (\mathbf{dM}_1 - \mathbf{dM}_2)\big]\mathbf{A}_1\mathbf{A}_1^{\mathrm{T}} + (\mathbf{M}_2 + \mathbf{dM}_2)(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{A}_1^{\mathrm{T}} \\
&\quad + (\mathbf{M}_2 + \mathbf{dM}_2)\mathbf{A}_2(\mathbf{A}_1 - \mathbf{A}_2)^{\mathrm{T}} - \mathbf{Y}_t(\mathbf{A}_1 - \mathbf{A}_2)^{\mathrm{T}}\Big\|_{\mathrm{F}}.
\end{aligned}$$

Since φ is Lipschitz continuous, there exists $\xi > 0$ such that

$$\begin{aligned}
\|\mathbf{A}_1 - \mathbf{A}_2\|_{\mathrm{F}} &\leq \|\varphi(\mathbf{M}_1) - \varphi(\mathbf{M}_2)\|_{\mathrm{F}} \leq \xi \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}} \\
\|\mathbf{dM}_1 - \mathbf{dM}_2\|_{\mathrm{F}} &\leq \|\varphi(\mathbf{M}_1) - \varphi(\mathbf{M}_2)\|_{\mathrm{F}} \leq \xi \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}.
\end{aligned}$$

In addition, $\varphi(\mathbf{M}) \in \mathcal{Z}_t$. Thus, there exists $\kappa > 0$ such that

$$\left\|\nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big) - \nabla_{\mathbf{M}} f\big(\mathbf{Y}_t, \mathbf{M}_2, \varphi(\mathbf{M}_2)\big)\right\|_{\mathrm{F}} \leq \kappa \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}$$

which concludes the proof in this first case.

Second case. If $(\mathbf{A}_t^k, \mathbf{dM}_t^k) \notin \operatorname{int} \mathcal{Z}_t$, the implicit function theorem can be applied to the auxiliary function $\tilde{f}$ defined as

$$\tilde{f}\big(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}\big) = f\big(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}\big) + \big\langle \mathbf{R}_t, (\mathbf{A}, \mathbf{dM})\big\rangle \tag{C.23}$$

where $\mathbf{R}_t = \big(-\nabla_{\mathbf{A}} f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k), -\nabla_{\mathbf{dM}} f(\mathbf{Y}_t, \tilde{\mathbf{M}}, \mathbf{A}_t^k, \mathbf{dM}_t^k)\big)$ is a constant. Similarly to the previous paragraph, there exists a neighborhood $V \times W$ of $\big(\tilde{\mathbf{M}}, (\mathbf{A}_t^k, \mathbf{dM}_t^k)\big)$ and a continuously differentiable function $\varphi : V \to W$ such that

(i) $\big(\tilde{\mathbf{M}}, (\mathbf{A}_t^k, \mathbf{dM}_t^k)\big) \in V \times W \subset \mathcal{M} \times \mathcal{Z}_t$;

(ii) for all $(\mathbf{M}, (\mathbf{A}, \mathbf{dM})) \in V \times W$, we have

$$[\nabla_{(\mathbf{A}, \mathbf{dM})} \tilde{f}(\mathbf{Y}_t, \mathbf{M}, \mathbf{A}, \mathbf{dM}) = \mathbf{0}] \Rightarrow [(\mathbf{A}, \mathbf{dM}) = \varphi(\mathbf{M})]; \tag{C.24}$$

(iii)

$$\mathbf{J}_{\varphi}(\mathbf{Y}_t, \tilde{\mathbf{M}}) = -\underbrace{\mathbf{H}^{-1}_{(\mathbf{A}, \mathbf{dM})} \tilde{f}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[(L+N)R \times (L+N)R]} \underbrace{\mathbf{J}_{\tilde{f}}^{\mathbf{M}, (\mathbf{A}, \mathbf{dM})}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[(L+N)R \times LR]}. \tag{C.25}$$

Consequently, the function $\tilde{v}(\mathbf{Y}_t, \cdot) = \tilde{f}\big(\mathbf{Y}_t, \cdot, \varphi(\cdot)\big)$ is $\mathcal{C}^1$ on V, and

$$\underbrace{\mathbf{J}_{\tilde{v}}^{\mathbf{M}}(\mathbf{Y}_t, \tilde{\mathbf{M}})}_{[1 \times LR]} = \mathbf{J}_{\tilde{f}}^{\mathbf{M}}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big) + \underbrace{\mathbf{J}_{\tilde{f}}^{(\mathbf{A}, \mathbf{dM})}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)}_{[1 \times (L+N)R]} \underbrace{\mathbf{J}_{\varphi}(\mathbf{Y}_t, \tilde{\mathbf{M}})}_{[(L+N)R \times LR]}. \tag{C.26}$$

The differentiability of $v(\mathbf{Y}_t, \cdot) = \tilde{v}(\mathbf{Y}_t, \cdot) - \big\langle \mathbf{R}, \varphi(\cdot) \big\rangle$ on V directly follows, with

$$\mathbf{J}_v^{\mathbf{M}}(\mathbf{Y}_t, \tilde{\mathbf{M}}) = \mathbf{J}_{\tilde{v}}^{\mathbf{M}}(\mathbf{Y}_t, \tilde{\mathbf{M}}) - (\operatorname{vec} \mathbf{R})^{\mathrm{T}} \mathbf{J}_{\varphi}(\mathbf{Y}_t, \tilde{\mathbf{M}}). \tag{C.27}$$

As in the previous paragraph, the second claim is a consequence from the continuous differentiability of $v(\mathbf{Y}_t, \cdot)$, defined on a compact set.

Finally, the Lipschitz continuity of $\nabla_{\mathbf{M}} g$ on V results from the following observations.

- $\|\mathbf{J}_{\varphi}(\mathbf{Y}_t, \mathbf{M})\|_{\mathrm{F}}$ is bounded by a constant independent from $\mathbf{Y}_t$ and $\mathbf{M}$, which implies that φ is Lipschitz continuous on V;
- $\Big\|\mathbf{H}_{(\mathbf{A},\mathbf{dM})}^{-1} \tilde{f}\big(\mathbf{Y}_t, \tilde{\mathbf{M}}, \varphi(\tilde{\mathbf{M}})\big)\Big\|_{\mathrm{F}}$, $\|\mathbf{R}\|_{\mathrm{F}}$, $\|\varphi(\mathbf{M})\|_{\mathrm{F}}$ and $\Big\|\mathbf{J}_{\tilde{f}}^{(\mathbf{A},\mathbf{dM})}\Big\|_{\mathrm{F}}$ are upper-bounded by constants independent from $\mathbf{Y}_t$;
- Proceeding as in the first case, $\mathbf{J}_{\varphi}(\mathbf{Y}_t, \cdot)$ and $\mathbf{J}_f^{\mathbf{M},(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \cdot, \varphi(\cdot)\big)$ can be shown to be Lipschitz continuous on V. Introducing $\mathbf{M}_1, \mathbf{M}_2 \in V$, and $(\mathbf{A}_i, \mathbf{dM}_i) = \varphi(\mathbf{M}_i)$ for $i = 1, 2$, we have

$$\begin{aligned}
&\big\|\nabla_{\mathbf{dM}} \tilde{f}\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big) - \nabla_{\mathbf{dM}} \tilde{f}\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big)\big\|_{\mathrm{F}} \\
&= \big\|\big[(\mathbf{M}_1 - \mathbf{M}_2) + (\mathbf{dM}_1 - \mathbf{dM}_2)\big]\mathbf{A}_1\mathbf{A}_1^{\mathrm{T}} + (\mathbf{M}_2 + \mathbf{dM}_2)(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{A}_1^{\mathrm{T}} \\
&\quad + (\mathbf{M}_2 + \mathbf{dM}_2)\mathbf{A}_2(\mathbf{A}_1 - \mathbf{A}_2)^{\mathrm{T}} - \mathbf{Y}_t(\mathbf{A}_1 - \mathbf{A}_2)^{\mathrm{T}}\big\|_{\mathrm{F}}
\end{aligned}$$

$$\begin{aligned}
&\big\|\nabla_{\mathbf{A}} \tilde{f}\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big) - \nabla_{\mathbf{A}} \tilde{f}\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big)\big\|_{\mathrm{F}} \\
&= \big\|(\mathbf{M}_1 + \mathbf{dM}_1)^{\mathrm{T}}\big[(\mathbf{M}_1 + \mathbf{dM}_1)\mathbf{A}_1 - \mathbf{Y}_t\big] - (\mathbf{M}_2 + \mathbf{dM}_2)^{\mathrm{T}}\big[(\mathbf{M}_2 + \mathbf{dM}_2)\mathbf{A}_2 - \mathbf{Y}_t\big]\big\|_{\mathrm{F}} \\
&= \big\|\big[(\mathbf{M}_1 - \mathbf{M}_2) + (\mathbf{dM}_1 - \mathbf{dM}_2)\big]^{\mathrm{T}}\mathbf{Y}_t + (\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 + \mathbf{dM}_1^{\mathrm{T}}\mathbf{dM}_1)(\mathbf{A}_1 - \mathbf{A}_2) + \big[\mathbf{M}_1^{\mathrm{T}}(\mathbf{M}_1 - \mathbf{M}_2) \\
&\quad + (\mathbf{M}_1 - \mathbf{M}_2)^{\mathrm{T}}\mathbf{M}_2 + \mathbf{dM}_1^{\mathrm{T}}(\mathbf{dM}_1 - \mathbf{dM}_2) + (\mathbf{dM}_1 - \mathbf{dM}_2)^{\mathrm{T}}\mathbf{dM}_2\big]\mathbf{A}_2\big\|_{\mathrm{F}}
\end{aligned}$$

$$\begin{aligned}
&\Big\|\mathbf{J}_{\tilde{f}}^{\mathbf{M},(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \mathbf{M}_1, \varphi(\mathbf{M}_1)\big) - \mathbf{J}_{\tilde{f}}^{\mathbf{M},(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \mathbf{M}_2, \varphi(\mathbf{M}_2)\big)\Big\|_{\mathrm{F}} \\
&= \Big\|\Big\{\mathbf{I}_R \otimes \big[(\mathbf{M}_1 + \mathbf{dM}_1)\mathbf{A}_1 - (\mathbf{M}_2 + \mathbf{dM}_2)\mathbf{A}_2\big]\Big\}\mathbf{S}_{R,N} + \big[\mathbf{A}_1 \otimes (\mathbf{M}_1 + \mathbf{dM}_1)\big] - \big[\mathbf{A}_2 \otimes (\mathbf{M}_2 + \mathbf{dM}_2)\big]\Big\|_{\mathrm{F}} \\
&= \Big\|\Big\{\mathbf{I}_R \otimes \big[(\mathbf{M}_1 - \mathbf{M}_2)\mathbf{A}_1 + (\mathbf{dM}_1 - \mathbf{dM}_2)\mathbf{A}_1 + (\mathbf{A}_1 - \mathbf{A}_2)(\mathbf{M}_2 + \mathbf{dM}_2)\big]\Big\} \\
&\quad + \big[\mathbf{A}_1 - \mathbf{A}_2 \otimes (\mathbf{M}_1 + \mathbf{dM}_1)\big] + \big[\mathbf{A}_2 \otimes (\mathbf{M}_1 - \mathbf{M}_2)\big] + \big[\mathbf{A}_2 \otimes (\mathbf{dM}_1 - \mathbf{dM}_2)\big]\Big\}\Big\|.
\end{aligned}$$

These expressions, along with the observations from the previous lines, show that $\mathbf{J}_{\varphi}(\mathbf{Y}_t, \cdot)$ and $\mathbf{J}_f^{\mathbf{M},(\mathbf{A},\mathbf{dM})}\big(\mathbf{Y}_t, \cdot, \varphi(\cdot)\big)$ are Lipschitz continuous on V.

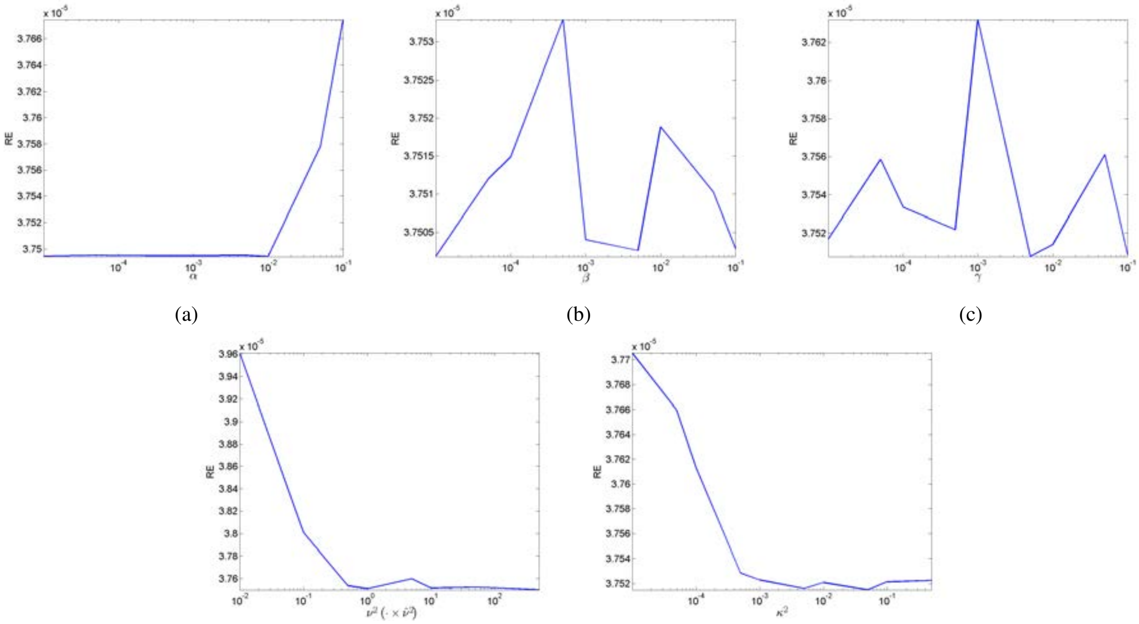
The preceding remarks finally lead to the conclusion that $\mathbf{J}_v^{\mathbf{M}}(\mathbf{Y}_t, \cdot)$ is Lipschitz continuous, which is sufficient to prove the third claim. □

C.4. Impact of the hyperparameters on the reconstruction error

Considering the significant number of hyper-parameters to be tuned (i.e., $\alpha, \beta, \gamma, \nu, \kappa$), 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., $(\alpha, \beta, \gamma, \nu^2, \kappa^2) = (10^{-2}, 10^{-4}, 10^{-4}, \hat{\nu}^2, 10^{-3})$, where $\hat{\nu}^2 = 0.0372$ 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 obtained solutions. The results reported in Fig. C.1 suggest that the proposed method is relatively robust to the choice of the hyper-parameters. More precisely, as can be seen in Figs. C.1(b) and C.1(c), 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 this section, the two following remarks can be made on the choice of ν and κ:

- the value chosen for ν results from an empirical compromise between the risk to capture noise into the variability terms (ν too large) and the risk to lose information (ν too small). The sensitivity analysis conducted in Fig. C.1(d) shows that $\nu^2 \in [10^{-1}, 1]$ 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. C.1(e) shows that $\kappa^2 \in [10^{-3}, 1]$ provides interesting results for the synthetic dataset used in the experiment.



(a) (b) (c)

(d) (e)

Figure C.1.: Sensitivity analysis of the reconstruction error RE with respect tothe tuning of the algorithm hyper-parameters ($\hat{\nu}^2 = 0.0372$ denotes the theoretical average energy of the variability introduced in the synthetic dataset used for this analysis).

APPENDIX D

Appendix to chapter 4

Algorithm convergence: sketch of proof

The proposed sketch of proof adapts the first arguments developed in [Can+16], in order to make clear that the proposed algorithm fits within this general framework. Note that a similar proof can be obtained by induction when J blocks have to be updated by each worker, and I blocks by the master node (corresponding to the situation described in (4.1)).

Lemma 2. *Under Assumptions 4 to 6, there exists two pisitive constants $c_{\mathbf{x}}$ and $c_{\mathbf{z}}$ such that*

$$\begin{aligned}\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1}) \leq \Psi(\mathbf{x}^{k},\mathbf{z}^{k}) - \frac{\gamma_k}{2}\big(c_{\mathbf{x}} - \gamma_k(L^+_{\mathbf{x}} + L^+_{\mathbf{x},\mathbf{z}})\big)\|\hat{\mathbf{x}}^k_{t^k} - \mathbf{x}^k_{t^k}\|^2 - \frac{\gamma_k}{2}\big(c_{\mathbf{z}} - \gamma_k L^+_{\mathbf{z}}\big)\|\hat{\mathbf{z}}^k - \mathbf{z}^k\|^2 \\ + \frac{1}{2}\tau(L^+_{\mathbf{x},\mathbf{z}} + L^+_{\mathbf{y},\mathbf{z}}) \sum_{q=k-\tau+1}^{k} \|\mathbf{z}^q - \mathbf{z}^{q-1}\|^2. \qquad \text{(D.1)}\end{aligned}$$

Proof. Assumption 5(v) allows the descent lemma [Ber99, p. 683] to be applied to $\mathbf{z} \mapsto F(\mathbf{x},\mathbf{y},\mathbf{z})$, leading to

$$F(\mathbf{x}^{k+1},\mathbf{z}^{k+1}) \leq F(\mathbf{x}^{k+1},\mathbf{z}^{k}) + \frac{L^k_{\mathbf{z}}}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \big\langle \nabla_{\mathbf{z}} F(\mathbf{x}^{k+1},\mathbf{z}^k), \mathbf{z}^{k+1} - \mathbf{z}^k\big\rangle. \qquad \text{(D.2)}$$

Thus,

$$\begin{aligned}\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1}) &\leq F(\mathbf{x}^{k+1},\mathbf{z}^{k}) + G(\mathbf{x}^{k+1}) + r(\mathbf{z}^{k+1}) + \big\langle \nabla_{\mathbf{z}} F(\mathbf{x}^{k+1},\mathbf{z}^k), \mathbf{z}^{k+1} - \mathbf{z}^k\big\rangle \\ &\quad + \frac{L^k_{\mathbf{z}}}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \\ &= f_{t^k}(\mathbf{x}^{k+1}_{t^k},\mathbf{z}^k) + g_{t^k}(\mathbf{x}^{k+1}_{t^k}) + \frac{L^k_{\mathbf{z}}}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2\big\langle \nabla_{\mathbf{z}} F(\mathbf{x}^{k+1},\mathbf{z}^k), \mathbf{z}^{k+1} - \mathbf{z}^k\big\rangle \\ &\quad + \sum_{t \neq t_k} f_t(\mathbf{x}^k_t,\mathbf{z}^k) + g_t(\mathbf{x}^k_t) + r(\mathbf{z}^{k+1}). \qquad \text{(D.3)}\end{aligned}$$

In addition, the optimality of $\hat{\mathbf{z}}^k$ implies

$$r(\hat{\mathbf{z}}^k) + \big\langle \nabla_{\mathbf{z}} F(\mathbf{x}^{k+1},\mathbf{z}^k), \hat{\mathbf{z}}^k - \mathbf{z}^k\big\rangle + \frac{c^k_{\mathbf{z}}}{2}\|\hat{\mathbf{z}}^k - \mathbf{z}^k\|^2 \leq r(\mathbf{z}^k) \qquad \text{(D.4)}$$

and the convexity of r leads to

$$r(\mathbf{z}^{k+1}) \leq r(\mathbf{z}^k) + \gamma_k \big(r(\hat{\mathbf{z}}^k) - r(\mathbf{z}^k) \big). \tag{D.5}$$

Combining (D.3), (D.4) and (D.5) then results in

$$\begin{aligned} \Psi(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \leq f_{t^k}(\mathbf{x}_{t^k}^{k+1}, \mathbf{z}^k) + g_{t^k}(\mathbf{x}_{t^k}^{k+1}) + r(\mathbf{z}^k) + \sum_{t \neq t_k} f_t(\mathbf{x}_t^k, \mathbf{z}^k) + g_t(\mathbf{x}_t^k) \\ - \frac{\gamma_k}{2}(c_{\mathbf{z}}^k - \gamma^k L_{\mathbf{z}}^k) \|\hat{\mathbf{z}}^k - \mathbf{z}^k\|^2. \end{aligned} \tag{D.6}$$

Similar arguments lead to

$$\begin{aligned} f_{t^k}(\mathbf{x}_{t^k}^{k+1}, \mathbf{z}^k) + g_{t^k}(\mathbf{x}_{t^k}^{k+1}) \leq f_{t^k}(\mathbf{x}_{t^k}^k, \mathbf{z}^k) - \frac{\gamma_k}{2}(c_{\mathbf{x}_{t^k}}^k - \gamma_k L_{\mathbf{x}_{t^k}}^k) \|\hat{\mathbf{x}}_{t^k}^k - \mathbf{x}_{t^k}^k\|^2 + g_{t^k}(\mathbf{x}_{t^k}^k) \\ + \big\langle \nabla_{\mathbf{x}_{t^k}} f_{t^k}(\mathbf{x}_{t^k}^k, \mathbf{z}^k) - \nabla_{\mathbf{x}_{t^k}} f_{t^k}(\mathbf{x}_{t^k}^k, \tilde{\mathbf{z}}^k), \mathbf{x}_{t^k}^{k+1} - \mathbf{x}_{t^k}^k \big\rangle. \end{aligned} \tag{D.7}$$

Since $\nabla_{\mathbf{x}_t} f_t(\mathbf{x}_t, \cdot)$ is assumed to be Lipschitz continuous (see Assumption 6(ii))

$$\big\langle \nabla_{\mathbf{x}_{t^k}} f_{t^k}(\mathbf{x}_{t^k}^k, \mathbf{z}^k) - \nabla_{\mathbf{x}_{t^k}} f_{t^k}(\mathbf{x}_{t^k}^k, \tilde{\mathbf{z}}^k), \mathbf{x}_{t^k}^{k+1} - \mathbf{x}_{t^k}^k \big\rangle \leq L_{\mathbf{x},\mathbf{z}}^k \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\| \|\mathbf{x}_{t^k}^{k+1} - \mathbf{x}_{t^k}^k\|$$

we obtain from (D.7)

$$\begin{aligned} f_{t^k}(\mathbf{x}_{t^k}^{k+1}, \mathbf{z}^k) + g_{t^k}(\mathbf{x}_{t^k}^{k+1}) \leq f_{t^k}(\mathbf{x}_{t^k}^k, \mathbf{z}^k) + L_{\mathbf{x},\mathbf{z}}^k \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\| \|\mathbf{x}_{t^k}^{k+1} - \mathbf{x}_{t^k}^k\| + g_{t^k}(\mathbf{x}_{t^k}^k) \\ - \frac{\gamma_k}{2}(c_{\mathbf{x}_{t^k}}^k - \gamma_k L_{\mathbf{x}_{t^k}}^k) \|\hat{\mathbf{x}}_{t^k}^k - \mathbf{x}_{t^k}^k\|^2. \end{aligned} \tag{D.8}$$

From this point, the product involving $\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|$ in (D.8) can be bounded as proposed in [Dav16, Proof of Theorem 5.1]. We first note that

$$L_{\mathbf{x},\mathbf{z}}^k \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\| \|\mathbf{x}_{t^k}^{k+1} - \mathbf{x}_{t^k}^k\| \leq \frac{L_{\mathbf{x},\mathbf{z}}^k}{2} \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|^2 + \frac{L_{\mathbf{x},\mathbf{z}}^k \gamma_k^2}{2} \|\hat{\mathbf{x}}_{t^k}^k - \mathbf{x}_{t^k}^k\|^2. \tag{D.9}$$

Besides, using Assumption 4

$$\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|^2 = \| \sum_{q=k-d_{t^k}^k+1}^{k} (\mathbf{z}^q - \mathbf{z}^{q-1}) \|^2 \leq \tau \sum_{q=k-\tau+1}^{k} \|\mathbf{z}^q - \mathbf{z}^{q-1}\|^2. \tag{D.10}$$

Combining (D.8), (D.9), and (D.10) then leads to

$$\begin{aligned} f_{t^k}(\mathbf{x}_{t^k}^{k+1}, \mathbf{z}^k) + g_{t^k}(\mathbf{x}_{t^k}^{k+1}) \leq f_{t^k}(\mathbf{x}_{t^k}^k, \mathbf{z}^k) - \frac{\gamma_k}{2}\big(c_{\mathbf{x}_{t^k}}^k - \gamma_k (L_{\mathbf{x}_{t^k}}^k + L_{\mathbf{x},\mathbf{z}}^k)\big) \|\hat{\mathbf{x}}_{t^k}^k - \mathbf{x}_{t^k}^k\|^2 \\ + \tau L_{\mathbf{x},\mathbf{z}}^k \sum_{q=k-\tau+1}^{k} \|\mathbf{z}^q - \mathbf{z}^{q-1}\|^2 + g_{t^k}(\mathbf{x}_{t^k}^k). \end{aligned} \tag{D.11}$$

Combining (D.6), (D.11) and using the bounds on the different Lipschitz constants introduced in Assumptions 5(vi) and 6(ii) finally leads to the announced result. ☐

According to Lemma 2, the objective function Ψ is not necessarily decreasing from an iteration to another due to the presence of a residual term involving τ past estimates of $\mathbf{z}$. From this observation, an auxiliary function (whose derivation is reproduced in Lemma 3 for the sake of completeness) has been proposed in [Dav16]. The introduction of such a function, which is eventually non-increasing between two consecutive iterations, is of a particular interest for the convergence analysis. This function finally allows convergence guarantees related to the original problem (4.2) to be recovered.

Lemma 3 (Auxiliary function definition, adapted from [Dav16, Proof of Theorem 5.1])**.** *Under the same assumptions as in Lemma 2, let Φ be the function defined by*

$$\Phi\big(\mathbf{x}(0),\mathbf{z}(0),\mathbf{z}(1),\dots,\mathbf{z}(\tau)\big) = \Psi\big(\mathbf{x}(0),\mathbf{z}(0)\big) + \frac{\beta}{2}\sum_{q=1}^{\tau}(\tau-q+1)\|\mathbf{z}(q)-\mathbf{z}(q-1)\|^2 \tag{D.12}$$

with

$$\beta = \tau(L^+_{\mathbf{x},\mathbf{z}} + L^+_{\mathbf{y},\mathbf{z}}).$$

Let $\mathbf{w}^k = (\mathbf{x}^k,\mathbf{z}^k,\check{\mathbf{z}}^k)$ and $\check{\mathbf{z}}^k = (\mathbf{z}^{k-1},\dots,\mathbf{z}^{k-\tau})$ for any iteration index $k\in\mathbb{N}$ (with the convention $\mathbf{z}^q = \mathbf{z}^0$ if $q<0$). Then,

$$\begin{aligned}\Phi(\mathbf{w}^{k+1}) \leq \Phi(\mathbf{w}^k) &- \frac{\gamma_k}{2}\big(c_{\mathbf{x}} - \gamma_k(L^+_{\mathbf{x}} + L^+_{\mathbf{x},\mathbf{z}})\big)\|\hat{\mathbf{x}}^k_{t^k} - \mathbf{x}^k_{t^k}\|^2\\ &- \frac{\gamma_k}{2}\big(c_{\mathbf{z}} - \gamma_k(L^+_{\mathbf{z}} + \tau^2 L^+_{\mathbf{x},\mathbf{z}})\big)\|\hat{\mathbf{z}}^k - \mathbf{z}^k\|^2.\end{aligned} \tag{D.13}$$

Proof. The expression of the auxiliary function proposed in [Dav16] results from the following decomposition of the residual term $\sum_{q=k-\tau+1}^{k}\|\mathbf{z}^q-\mathbf{z}^{q-1}\|^2$. Introducing the auxiliary variables

$$\alpha^k = \sum_{q=k-\tau+1}^{k}(q-k+\tau)\|\mathbf{z}^q-\mathbf{z}^{q-1}\|^2$$

we can note that

$$\alpha^k - \alpha^{k+1} = \sum_{q=k-\tau+1}^{k}\|\mathbf{z}^q-\mathbf{z}^{q-1}\|^2 - \tau\|\mathbf{z}^{k+1}-\mathbf{z}^k\|^2. \tag{D.14}$$

Thus, using the upper bound $L^k_{\mathbf{x},\mathbf{z}} \leq L^+_{\mathbf{x},\mathbf{z}}$ (Assumption 6(ii)) and replacing (D.14) in (D.1) yields

$$\begin{aligned}\Psi(\mathbf{x}^{k+1},\mathbf{z}^{k+1}) + \beta\alpha^{k+1} \leq \Psi(\mathbf{x}^k,\mathbf{z}^k) + \beta\alpha^k &- \frac{\gamma_k}{2}\big(c_{\mathbf{x}} - \gamma_k(L^+_{\mathbf{x}} + L^+_{\mathbf{x},\mathbf{z}})\big)\|\hat{\mathbf{x}}^k_{t^k} - \mathbf{x}^k_{t^k}\|^2\\ &- \frac{\gamma_k}{2}\big(c_{\mathbf{z}} - \gamma_k(L^+_{\mathbf{z}} + \tau^2 L^+_{\mathbf{x},\mathbf{z}})\big)\|\hat{\mathbf{z}}^k - \mathbf{z}^k\|^2.\end{aligned}$$

Observing that $\Phi(\mathbf{w}^k) = \Psi(\mathbf{x}^k,\mathbf{z}^k) + \alpha^k$ finally leads to the announced result. ☐

The previous lemma makes clear that the proposed algorithm can be studied as a special case of [Can+16]. The rest of the convergence analysis, which involves somewhat convoluted arguments, exactly follows [Can+16] up to minor notational modifications.

Bibliography

[ABF10] M. V. Alfonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. “Fast image recovery using variable splitting and constrained optimization”. In: *IEEE Trans. Image Process.* 19.9 (Sept. 2010), pp. 2345–2356 (cit. on p. 18).

[ABF11] M. V. Alfonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems”. In: *IEEE Trans. Image Process.* 20.3 (Mar. 2011), pp. 681–695 (cit. on p. 18).

[ADT11] 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 (cit. on p. 34).

[ADT14] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret. “Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm”. In: *IEEE Trans. Image Process.* 23.6 (June 2014), pp. 2663–2675 (cit. on pp. 7, 54).

[AF13a] 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 (cit. on pp. 18, 22).

[AF13b] M. S. C. Almeida and M. A. T. Figueiredo. “Deconvolving images with unknown boundaries using the alternating direction method of multipliers”. In: *arXiv:1210.2687v2* (Mar. 2013) (cit. on p. 18).

[Alt+12] Y. Altmann et al. “Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery”. In: *IEEE Trans. Image Process.* 21.6 (June 2012), pp. 3017–3025 (cit. on p. 7).

[Alt+13] Y. Altmann et al. “A robust test for nonlinear mixture detection in hyperspectral images”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Vancouver, Canada, June 2013, pp. 2149–2153 (cit. on pp. 68, 71).

[Alt+14] Y. Altmann et al. “Residual component analysis of hyperspectral images - Application to joint nonlinear unmixing and nonlinearity detection”. In: *IEEE Trans. Image Process.* 23.5 (June 2014), pp. 2148–2158 (cit. on p. 49).

[Alt13] Y. Altmann. “Nonlinear unmixing of hyperspectral images”. PhD thesis. Institut National Polytechnique de Toulouse, 2013. URL: https://tel.archives-ouvertes.fr/tel-00945513v1 (cit. on pp. 1, 5).

[AMD14] Y. Altmann, S. McLaughlin, and N. Dobigeon. “Sampling from a multivariate Gaussian distribution truncated on a simplex: a review”. In: *Proc. IEEE-SP Workshop Stat. and Signal Processing (SSP)*. Gold Coast, Australia, July 2014, pp. 113–116 (cit. on p. 54).

[AMH15] Y. Altmann, S. McLaughlin, and A. O. Hero. “Robust linear spectral unmixing using anomaly detection”. In: *IEEE Trans. Comput. Imag.* 1.2 (June 2015), pp. 74–85 (cit. on pp. 2, 13, 44–46, 49, 56, 143).

[Amm+17] R. Ammanouil et al. “Nonlinear unmixing of hyperspectral data with vector-valued kernel functions”. In: *IEEE Trans. Image Process.* 26.1 (Jan. 2017), pp. 340–354 (cit. on p. 7).

[ASL11] M. Arngren, M. N. Schmidt, and J. Larsen. “Unmixing of hyperspectral images using Bayesian nonnegative matrix factorization with volume prior”. In: *J. Signal Process. Sys.* 65.3 (Nov. 2011), pp. 479–496 (cit. on pp. 21, 80).

[BC05] S. Bourguignon and H. Carfantan. “Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data”. In: *Proc. IEEE-SP Workshop Stat. and Signal Processing (SSP)*. Bordeaux, France, July 2005, pp. 811–816 (cit. on p. 49).

[BCG11] S. Becker, E. J. Candès, and M. Grant. “Templates for convex cone problems with applications to sparse signal recovery”. In: *Math. Prog. Comp.* 3.3 (Sept. 2011), pp. 165–218 (cit. on p. 82).

[BCK14] J. Bioucas-Dias, F. Condessa, and J. Kovacevic. “Alternating direction optimization for image segmentation using hidden markov measure field models”. In: *Proc. SPIE 9019, Image Processing: Algorithms and Systems XII*. Feb. 2014, 90190P (cit. on pp. 72, 128, 132).

[BD86] 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 (cit. on pp. 82, 149).

[BDT11] C. Bazot, N. Dobigeon, and J.-Y. Tourneret. “Bernoulli-Gaussian model for gene expression analysis”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Prague, Czech Republic, May 2011, pp. 5996–5999 (cit. on p. 49).

[Ber+04] M. Berman et al. “ICE: a statistical approach to identifying endmembers in hyperspectral images”. In: *IEEE Trans. Geosci. Remote Sens.* 42.10 (Oct. 2004), pp. 2085–2095 (cit. on pp. 21, 80).

[Ber99] D. P. Bertsekas. *Nonlinear programming*. Belmont, MA: Athena Scientific, 1999 (cit. on pp. 25, 157).

[Bez+17] J. Bezanson et al. “Julia: a fresh approach to numerical computing”. In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671 (cit. on p. 109).

[BF10] 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 (cit. on pp. 10, 18, 29, 45, 46, 57, 58, 76, 77, 85, 109).

[BF16] J. M. Bioucas-Dias and M. A. T. Figueiredo. “Bayesian image segmentation using hidden fields: supervised, unsupervised, and semi-supervised formulations”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Budapest, Hungary, Aug. 2016, pp. 523–527 (cit. on pp. 72, 128, 132).

[BHI16] P. Bianchi, W. Hachem, and F. Iutzeler. “A Coordinate Descent Primal-Dual Algorithm and Application to Distributed Asynchronous Optimization”. In: *IEEE Trans. Autom. Control* 61.10 (Oct. 2016), p. 2947 (cit. on pp. 100, 115, 123, 128, 132).

[Bio+12] J. M. Bioucas-Dias et al. “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 5.2 (Apr. 2012), pp. 354–379 (cit. on pp. 1, 5–7).

[Bio09] 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 (cit. on pp. 18, 29, 45, 46, 57, 58, 76, 77, 85, 109, 111, 118).

[BJ13] P. Bianchi and J. Jakubowicz. “Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization”. In: *IEEE Trans. Autom. Control* 58.2 (Feb. 2013), pp. 391–405 (cit. on pp. 99, 100).

[BN08] J. M. Bioucas-Dias and J. M. P. Nascimento. “Hyperspectral subspace identification”. In: *IEEE Trans. Geosci. Remote Sens.* 46.8 (Aug. 2008), pp. 2435–2445 (cit. on pp. 6, 7, 91).

[Bob+07] J. Bobin et al. “Sparsity and morphological diversity in blind source separation”. In: *IEEE Trans. Image Process.* 16.11 (Nov. 2007), pp. 2662–2674 (cit. on pp. 127, 131).

[Bot16] Z. I. Botev. “The normal law under linear restrictions: simulation and estimation via minimax tilting”. In: *J. Roy. Stat. Soc. Ser. B* 78.3 (2016). to appear (cit. on p. 54).

[Boy+10] S. Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. In: *Machine Learning* 3.1 (2010), pp. 1–122 (cit. on pp. 2, 13, 25, 101, 136, 137).

[BST13] J. Bolte, S. Sabach, and M. Teboulle. “Proximal alternating linearized minimization for nonconvex and nonsmooth problems”. In: *Mathematical Programming* 1-2.146 (July 2013), pp. 459–494 (cit. on pp. 2, 3, 13, 14, 18, 25, 27, 41, 42, 76, 77, 81–83, 85, 86, 101, 104–108, 123, 125, 128, 140, 149).

[Can+09] E. J. Candès et al. “Robust principal component analysis?” In: *journaltitle of ACM* 58.1 (2009), pp. 1–37 (cit. on p. 50).

[Can+16] L. Cannelli et al. “Asynchronous Parallel Algorithms for Nonconvex Big-Data Optimization. Part I: Model and Convergence”. In: (July 2016). arXiv preprint. URL: `https://arxiv.org/abs/1607.04818` (cit. on pp. 2, 14, 100–105, 123, 157, 160).

[Cav+17] Y. C. Cavalcanti et al. “Unmixing dynamic PET images with a PALM algorithm”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. to appear. Kos, Greece, Sept. 2017 (cit. on pp. 128, 132).

[CB17] C. Chenot and J. Bobin. “Blind separation of sparse sources in the presence of outliers”. In: *Signal Processing* 138 (Sept. 2017), pp. 233–243 (cit. on pp. 128, 132).

[CBR15] C. Chenot, J. Bobin, and J. Rapin. “Robust sparse blind source separation”. In: *IEEE Signal Process. Lett.* 22.11 (Nov. 2015), pp. 2172–2176 (cit. on pp. 44, 45, 128, 132).

[CE16] P. L. Combettes and J. Eckstein. “Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions”. In: *Math. Program., Ser. B* (2016), pp. 1–28 (cit. on pp. 2, 99–101).

[Cha+11] T.-H. Chan et al. “A simplex volume maximization framework for hyperspectral endmember extraction”. In: *IEEE Trans. Geosci. Remote Sens.* 49.11 (Nov. 2011), pp. 4177–4193 (cit. on p. 21).

[Cha+16] T.-H. Chang et al. “Asynchronous Distributed ADMM for Large-Scale Optimization–Part I: Algorithm and Convergence Analysis”. In: *IEEE Trans. Signal Process.* 62.12 (June 2016), pp. 3118–3130 (cit. on pp. 2, 14, 101–103, 123).

[Cho11] N. Chopin. “Fast simulation of truncated Gaussian distributions”. In: *Statistics and Computing* 21.2 (2011), pp. 275–288. ISSN: 1573-1375 (cit. on p. 54).

[Con15] L. Condat. “Fast projection onto the simplex and the ℓ_1 ball”. In: *Math. Program., Ser. A* (Sept. 2015), pp. 1–11. DOI: `10.1007/s10107-015-0946-6`. URL: `http://dx.doi.org/10.1007/s10107-015-0946-6` (cit. on pp. 82, 108).

[CP11] P. L. Combettes and J.-C. Pesquet. “Proximal splitting methods in signal processing”. In: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* 49 (May 2011), pp. 185–212 (cit. on pp. 127, 131).

[CPR14] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. “Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function”. In: *J. Optim. Theory Appl.* 162.1 (2014), pp. 107–132 (cit. on pp. 123, 128, 132).

[CPR16] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. “A block coordinate variable metric forward-backward algorithm”. In: *J. Glob. Optim* 66.3 (2016), pp. 547–485 (cit. on pp. 2, 3, 11, 14, 18, 41, 42, 76, 77, 84, 102, 123, 125, 128, 132).

[CRH13] J. Chen, C. Richard, and P. Honeine. “Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model”. In: *IEEE Trans. Signal Process.* 61.2 (Jan. 2013), pp. 480–492 (cit. on p. 7).

[CRH14] J. Chen, C. Richard, and P. Honeine. “Nonlinear estimation of material abundance in hyperspectral images with ℓ_1-norm spatial regularization”. In: *IEEE Trans. Geosci. Remote Sens.* 52.5 (May 2014), pp. 2654–2665 (cit. on p. 21).

[CTC15] L. Chaari, J.-Y. Tourneret, and C. Chaux. “Sparse signal recovery using a Bernoulli generalized Gaussian prior”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Nice, France, Sept. 2015, pp. 1711–1715 (cit. on p. 49).

[CTS15] J. Chen, Z. J. Towfic, and A. H. Sayed. “Dictionary learning over distributed models”. In: *IEEE Trans. Signal Process.* 63.4 (Feb. 2015), pp. 1001–1016 (cit. on pp. 99, 100).

[Dav16] D. Davis. “The asynchronous PALM algorithm for nonsmooth nonconvex problems”. In: (Apr. 2016). submitted. URL: `https://arxiv.org/abs/1604.00526` (cit. on pp. 2, 14, 100–103, 106, 109, 111, 117, 118, 120, 121, 123, 158, 159).

[DCJ16] L. Drumetz, J. Chanussot, and C. Jutten. “Variability of the endmembers in spectral unmixing: recent advances”. In: *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*. Los Angeles, United States, Aug. 2016. URL: `https://hal.archives-ouvertes.fr/hal-01358368` (cit. on pp. 8, 10).

[Deb+14] C. Debes et al. “Hyperspectral and lidar data fusion: outcome of the 2013 grss data fusion contest”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 7.6 (June 2014), pp. 2405–2418 (cit. on p. 116).

[Del+14] C.-A. Deledalle et al. “Stein unbiased gradient estimator of the risk (sugar) for multiple parameter selection”. In: *SIAM J. Imaging Sciences* 7.4 (2014), pp. 2448–2487 (cit. on pp. 127, 131).

[DHT09] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret. “Hierarchical Bayesian sparse image reconstruction with application to MRFM”. In: *IEEE Trans. Image Process.* 18.9 (Sept. 2009), pp. 2059–2070 (cit. on p. 49).

[Dob+09] N. Dobigeon et al. “Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery”. In: *IEEE Trans. Signal Process.* 57.11 (Nov. 2009), pp. 4355–4368 (cit. on pp. 34, 50, 135).

[Dob+14a] N. Dobigeon et al. “A comparison of nonlinear mixing models for vegetated areas using simulated and real hyperspectral data”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 7.6 (June 2014), pp. 1869–1878. URL: `http://arxiv.org/abs/1312.1270/` (cit. on p. 7).

[Dob+14b] N. Dobigeon et al. “Nonlinear unmixing of hyperspectral images: models and algorithms”. In: *IEEE Signal Process. Mag.* 31.1 (Jan. 2014), pp. 89–94 (cit. on pp. 1, 5, 7, 34).

[Dob07] N. Dobigeon. “Modèles bayésiens hiérarchiques pour le traitement multi-capteur”. PhD thesis. Institut National Polytechnique de Toulouse, 2007. URL: `https://tel.archives-ouvertes.fr/tel-00189738v1` (cit. on pp. 1, 5).

[Dru+16] L. Drumetz et al. “Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability”. In: *IEEE Trans. Image Process.* 25.8 (Aug. 2016), pp. 3890–3905 (cit. on pp. 1, 5, 9, 10, 18–20, 29, 30, 33, 37, 39, 59, 125, 127, 131).

[DT07] N. Dobigeon and J.-Y. Tourneret. *Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex*. Tech. rep. France: IRIT/ENSEEIHT/TéSA, Mar. 2007. URL: `http://www.enseeiht.fr/~dobigeon/papers/Dobigeon_TechReport_2007b.pdf` (cit. on p. 54).

[Du+14] X. Du et al. “Spatial and spectral unmixing using the beta compositional model”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 7.6 (June 2014), pp. 1994–2003 (cit. on pp. 1, 5, 8, 9).

[Duc+08] J. Duchi et al. “Efficient projection onto the ℓ_1-ball for learning in high dimensions”. In: *Proc. Int. Conf. Machine Learning (ICML)*. Helsinki, Finland, 2008 (cit. on pp. 82, 108).

[Ech+10] O. Eches et al. “Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery”. In: *IEEE Trans. Image Process.* 19.6 (June 2010), pp. 1403–1413 (cit. on pp. 1, 5, 8, 9, 20).

[Ech10] O. Eches. “Méthodes Bayésiennes pour le démélange d’images hyperspectrales”. PhD thesis. Institut National Polytechnique de Toulouse, Univ. of Toulouse, 2010. URL: `http://ethesis.inp-toulouse.fr/archive/00001309/` (cit. on pp. 1, 5).

[EDT11] O. Eches, N. Dobigeon, and J. Y. Tourneret. “Enhancing hyperspectral image unmixing with spatial correlations”. In: *IEEE Trans. Geosci. Remote Sens.* 49.11 (Nov. 2011), pp. 4239–4247 (cit. on pp. 34, 49).

[EP15] A. Ertürk and A. Plaza. “Informative Change Detection by Unmixing for Hyperspectral Images”. In: *IEEE Geosci. Remote Sens. Lett.* 12.6 (June 2015), pp. 1252–1256 (cit. on pp. 11, 44, 45).

[Fan+09] W. Fan et al. “Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data”. In: *Int. J. Remote Sens.* 30.11 (June 2009), pp. 2951–2962 (cit. on p. 7).

[FD15] C. Févotte and N. Dobigeon. “Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization”. In: *IEEE Trans. Image Process.* 24.12 (Dec. 2015), pp. 4904–4917 (cit. on pp. 34, 45, 46, 57, 59).

[Fer+17] V. Ferraris et al. “Robust fusion of multiband images with different spatial and spectral resolutions for change detection”. In: *IEEE Trans. Comput. Imag.* 3.2 (June 2017), pp. 175–186 (cit. on pp. 128, 132).

[FGP15] P. Frankel, G. Garrigos, and J. Peypouquet. “Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates”. In: *J. Optim. Theory Appl.* 165.1 (2015), pp. 874–900 (cit. on pp. 102, 123, 128, 132).

[Fre+17] J. Frecon et al. “Bayesian selection for the l2-potts model regularization parameter: 1d piecewise constant signal denoising”. In: *IEEE Trans. Signal Process.* (2017). to appear (cit. on pp. 128, 131).

[FSS15] F. Facchinei, G. Scutari, and S. Sagratella. “Parallel selective algorithms for nonconvex big data optimization”. In: *IEEE Trans. Signal Process.* 63.7 (Apr. 2015), pp. 1874–1889 (cit. on pp. 99, 100).

[Gad+13] P. Gader et al. *MUUFL gulfport hyperspectral and LiDAR airborne data set*. Tech. rep. REP-2013-570. Gainesville, FL: University of Florida, Oct. 2013 (cit. on p. 8).

[Gio10] J.-F. Giovannelli. “Estimation of the Ising field parameter thanks to the exact partition function”. In: *Proc. IEEE Int. Conf. Image Processing (ICIP)*. Hong-Kong, China, Sept. 2010, pp. 1441–1444 (cit. on pp. 49, 56).

[Gio11] J.-F. Giovannelli. “Ising field parameter estimation from incomplete and noisy data”. In: *Proc. IEEE Int. Conf. Image Processing (ICIP)*. Brussels, Belgium, Sept. 2011, pp. 1853–1856 (cit. on p. 52).

[GL80] G. H. Golub and C. F. V. Loan. “An analysis of the total least squares problem”. In: *SIAM J. Numer. Anal.* 17.6 (Dec. 1980), pp. 883–893 (cit. on pp. 2, 13, 17–19, 129).

[Goe+13] M. A. Goenaga et al. “Unmixing analysis of a time series of hyperion images over the Guánica dry forest in Puerto Rico”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 6.2 (Apr. 2013), pp. 329–338 (cit. on pp. 11, 29, 44, 45, 76).

[GR92] A. Gelman and D. B. Rubin. “Inference from iterative simulation using multiple sequences”. In: *Statistical Science* 7.4 (1992), pp. 457–511 (cit. on p. 143).

[Hal+11] A. Halimi et al. “Nonlinear unmixing of hyperspectral images using a generalized bilinear model”. In: *IEEE Trans. Geosci. Remote Sens.* 49.11 (Nov. 2011), pp. 4153–4162 (cit. on pp. 7, 34).

[Hal+15] A. Halimi et al. “Unmixing hyperspectral images accounting for temporal and spatial endmember variability”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Nice, France, Sept. 2015, pp. 1686–1690 (cit. on pp. 1, 2, 5, 11, 13, 44, 45, 76).

[Hal+16a] A. Halimi et al. “Bayesian estimation of smooth altimetric parameters: application to conventional and delay/Doppler altimetry”. In: *IEEE Trans. Geosci. Remote Sens.* 54.4 (Apr. 2016), pp. 2207–2219 (cit. on pp. 45, 75, 76).

[Hal+16b] A. Halimi et al. “Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach”. In: *IEEE Trans. Geosci. Remote Sens.* 54.7 (July 2016), pp. 3811–3821 (cit. on pp. 6, 7, 66, 91, 116).

[Hal+17] A. Halimi et al. “Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects”. In: *IEEE Trans. Comput. Imag.* 3.2 (June 2017), pp. 146–159 (cit. on pp. 127, 131).

[Hap81] B. W. Hapke. “Bidirectional reflectance spectroscopy. I. Theory”. In: *J. Geophys. Res.* 86.B4 (Apr. 1981), pp. 3039–3054 (cit. on p. 10).

[Hap93] B. Hapke. *Theory of reflectance and emittance spectroscopy*. Cambridge, UK: Cambridge Univ. Press, 1993 (cit. on p. 10).

[HAS13] R. Heylen, A. Atker, and P. Scheunders. “On using projection onto convex sets for solving the hyperspectral unmixing problem”. In: *IEEE Geosci. Remote Sens. Lett.* 10.6 (2013), pp. 1522–1526 (cit. on pp. 82, 149).

[HC00] D. C. Heinz and C.-I Chang. “Constrained subpixel target detection for remotely sensed imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 38 (May 2000), pp. 1144–1159 (cit. on p. 29).

[HC01] D. C. Heinz and C. -I Chang. “Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 29.3 (Mar. 2001), pp. 529–545 (cit. on pp. 76, 77).

[HCJ16] S. Henrot, J. Chanussot, and C. Jutten. “Dynamical spectral unmixing of multitemporal hyperspectral images”. In: *IEEE Trans. Image Process.* 25.7 (July 2016), pp. 3219–3232 (cit. on pp. 1, 2, 5, 11, 44–46, 59, 76, 77, 85, 86, 88, 108, 109, 116).

[HDT15] A. Halimi, N. Dobigeon, and J.-Y. Tourneret. “Unsupervised unmixing of hyperspectral images accounting for endmember variability”. In: *IEEE Trans. Image Process.* 24.12 (Dec. 2015), pp. 4904–4917 (cit. on pp. 1, 5, 8, 9, 20).

[Hey+16] R. Heylen et al. “Hyperspectral unmixing with endmember variability via alternating angle minimization”. In: *IEEE Trans. Geosci. Remote Sens.* 54.8 (Aug. 2016), pp. 4983–4993 (cit. on p. 10).

[HMT15] A. Halimi, C. Mailhes, and J.-Y. Tourneret. “Nonlinear regression using smooth Bayesian estimation”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Brisbane, Australia, Apr. 2015, pp. 2634–2638 (cit. on pp. 45, 48, 49, 75, 76).

[HPG14] R. Heylen, M. Parente, and P. Gader. “A review of nonlinear hyperspectral unmixing methods”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 7.6 (June 2014), pp. 1844–1868 (cit. on pp. 1, 5, 7).

[HY16] R. Hannah and W. Yin. “On unbounded delays in asynchronous parallel fixed-point algorithms”. In: *arXiv preprint* (Sept. 2016). URL: `https://arxiv.org/abs/1609.04746` (cit. on pp. 2, 14).

[IBP11] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza. “Sparse unmixing of hyperspectral data”. In: *IEEE Trans. Geosci. Remote Sens.* 49.6 (June 2011), pp. 2014–2039 (cit. on p. 10).

[JJ13] 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 (cit. on pp. 17, 18).

[JWZ10] J. Jin, B. Wang, and L. Zhang. “A novel approach based on Fisher discriminant null space for decomposition of mixed pixels in hyperspectral imagery”. In: *IEEE Geosci. Remote Sens. Lett.* 7.4 (Oct. 2010), pp. 699–703 (cit. on pp. 1, 5, 10, 18, 29).

[KM82] J. J. Kormylo and J. M. Mendel. “Maximum likelihood detection and estimation of Bernoulli-Gaussian processes”. In: *IEEE Trans. Inf. Theory* 28.3 (May 1982), pp. 482–488 (cit. on p. 49).

[Lan99] S. Lang. *Fundamentals of differential geometry*. 1st ed. Vol. 191. Graduate Texts in Mathematics. Springer-Verlag New York, 1999. DOI: `10.1007/978-1-4612-0541-8` (cit. on p. 152).

[Lav93] M. Lavielle. “Bayesian deconvolution of Bernoulli-Gaussian processes”. In: *Signal Process.* 33.1 (July 1993), pp. 67–79 (cit. on p. 49).

[Lia+14] J. Liang et al. “Distributed dictionary learning for sparse representation in sensor networks”. In: *IEEE Trans. Image Process.* 23.6 (June 2014), pp. 2528–2541 (cit. on pp. 99, 100).

[Lia+15] X. Lian et al. “Asynchronous parallel stochastic gradient for nonconvex optimization”. In: *Adv. in Neural Information Processing Systems*. Montréal, Canada, Dec. 2015, pp. 2719–2727 (cit. on p. 101).

[Liu+16] S. Liu et al. “Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images”. In: *IEEE Trans. Geosci. Remote Sens.* 54.5 (May 2016), pp. 2733–2748 (cit. on pp. 11, 44, 45).

[LS15] P. D. Lorenzo and G. Scutari. “Distributed nonconvex optimization over networks”. In: *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*. Cancun, Mexico, Dec. 2015, pp. 229–232 (cit. on pp. 99, 100).

[Mai+10] J. Mairal et al. “Online learning for matrix factorization and sparse coding”. In: *J. Mach. Learning Research* 11 (Jan. 2010), pp. 19–60 (cit. on pp. 2, 13, 76, 79, 81–84, 86, 150, 151).

[Maz+15] V. Mazet et al. “Unsupervised joint decomposition of a spectroscopic signal sequence”. In: *Signal Process.* 109 (Apr. 2015), pp. 193–205 (cit. on p. 48).

[Meg+14] I. Meganem et al. “Linear-quadratic mixing model for reflectances in urban environments”. In: *IEEE Trans. Geosci. Remote Sens.* 52.1 (Jan. 2014), pp. 544–558 (cit. on p. 7).

[Mey+16] T. Meyer et al. “Hyperspectral unmixing with material variability using social sparsity”. In: *Proc. IEEE Int. Conf. Image Processing (ICIP)*. Pheonix, United States, Sept. 2016 (cit. on p. 10).

[MQ07] L. Miao and H. Qi. “Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization”. In: *IEEE Trans. Geosci. Remote Sens.* 45.3 (Mar. 2007), pp. 765–776 (cit. on pp. 21, 34, 116, 117).

[MRF13] A. Matakos, S. Ramani, and J. A. Fessler. “Accelerated edge-preserving image restoration with boundary artifacts”. In: *IEEE Trans. Image Process.* 22.5 (May 2013), pp. 2019–2029 (cit. on p. 18).

[MSB03] J. L. Marroquin, E. A. Santana, and S. Botello. “Hidden Markov measure field models for image segmentation”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 25.11 (2003), pp. 1380–1387 (cit. on pp. 72, 128, 132).

[NB05a] J. M. P. Nascimento and J. M. Bioucas-Dias. “Does independent component analysis play a role in unmixing hyperspectral data ?” In: *IEEE Trans. Geosci. Remote Sens.* 43.1 (Jan. 2005), pp. 175–187 (cit. on p. 10).

[NB05b] J. M. Nascimento and J. M. Bioucas-Dias. “Vertex component analysis: a fast algorithm to unmix hyperspectral data”. In: *IEEE Trans. Geosci. Remote Sens.* 43.4 (Apr. 2005), pp. 898–910 (cit. on pp. 10, 18, 29, 34, 38–40, 45, 46, 57, 58, 76, 77, 85, 109, 111, 115–118).

[NB09] J. M. P. Nascimento and J. M. Bioucas-Dias. “Nonlinear mixture model for hyperspectral unmixing”. In: *Proc. SPIE Image and Signal Processing for Remote Sensing XV*. Ed. by L. Bruzzone, C. Notarnicola, and F. Posa. Vol. 7477. 1. Berlin, Germany: SPIE, 2009, p. 74770I (cit. on p. 7).

[Ons44] L. Onsager. “A two-dimensional model with an order-disorder transition”. In: *Phys. Rev.* 65.3 & 4 (Feb. 1944), pp. 117–149 (cit. on p. 56).

[PB14] N. Parikh and S. Boyd. “Proximal algorithms”. In: *Foundations and Trends® in Optimization* 1.3 (2014), pp. 127–239. DOI: `10.1561/2400000003` (cit. on p. 149).

[PBF15] M. Pereyra, J. M. Bioucas-Dias, and M. A. T. Figueiredo. “Maximum-a-posteriori estimation with unknown regularisation parameters”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Nice, France, Aug. 2015, pp. 230–2343 (cit. on pp. 127, 131).

[Pen+16] Z. Peng et al. “Arock: an algorithmic framework for asynchronous parallel coordinate updates”. In: *SIAM J. Sci. Comput.* 38.5 (Sept. 2016), pp. 2851–2879 (cit. on pp. 2, 100, 101).

[Per+13] M. Pereyra et al. “Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm”. In: *IEEE Trans. Image Process.* 22.6 (June 2013), pp. 2385–2397 (cit. on p. 50).

[Per+14] M. Pereyra et al. “Computing the Cramer-Rao bound of Markov random field parameters: application to the Ising and the Potts models”. In: *IEEE Signal Process. Lett.* 21.1 (Jan. 2014), pp. 47–50 (cit. on p. 143).

[PP14] A. Pakman and L. Paninski. “Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians”. In: *journaltitle of Computational and Graphical Statistics* 23.2 (2014), pp. 518–542. eprint: `http://dx.doi.org/10.1080/10618600.2013.788448` (cit. on p. 54).

[PR15] J.-C. Pesquet and A. Repetti. “A Class of Randomized Primal-Dual Algorithms for Distributed Optimization”. In: *Journal of nonlinear and convex analysis* 16.12 (Nov. 2015), pp. 2453–2490 (cit. on pp. 2, 14, 99, 100, 115, 123, 128, 132).

[Pre+16] J. Prendes et al. “A Bayesian Nonparametric Model Coupled with a Markov Random Field for Change Detection in Heterogeneous Remote Sensing Images”. In: *SIAM J. Imaging Sciences* 9.4 (2016), pp. 1889–1921 (cit. on pp. 128, 132).

[Qia+11] Y. Qian et al. “Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization”. In: *IEEE Trans. Geosci. Remote Sens.* 49.11 (Nov. 2011), pp. 4282–4297 (cit. on pp. 76, 77, 85).

[Rap+14] J. Rapin et al. “NMF with sparse regularizations in transformed domains”. In: *SIAM J. Imaging Sciences* 7.4 (2014), pp. 2020–2047 (cit. on pp. 127, 128, 131, 132).

[RCP14] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. “A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Florence, Italy, May 2014, pp. 1498–1502 (cit. on pp. 123, 128, 132).

[Rob+98] D. Roberts et al. “Mapping chaparral in the Santa Monica mountain using multiple endmember spectral mixture models”. In: *Remote Sens. Environment* 65.3 (Sept. 1998), pp. 267–279 (cit. on pp. 9, 10, 29).

[RX11] D. Ralph and H. Xu. “Convergence of stationary points of sample average two-stage stochastic programs: a generalized equation approach”. In: *Mathematics of Operations Research* 36.3 (Aug. 2011), pp. 568–592 (cit. on pp. 2, 13, 76, 80, 81).

[Scu+17] G. Scutari et al. “Parallel and distributed methods for nonconvex optimization-part i: theory”. In: *IEEE Trans. Signal Process.* 65.8 (Apr. 2017), pp. 1929–2944 (cit. on pp. 2, 14, 99, 100).

[Som+11] B. Somers et al. “Endmember variability in spectral mixture analysis: a review”. In: *Remote Sens. Environment* 115.7 (July 2011), pp. 1603–1616 (cit. on pp. 7, 10).

[Som+12] B. Somers et al. “Automated extraction of image-based endmember bundles for improved spectral unmixing”. In: *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.* 5.2 (Apr. 2012), pp. 396–408 (cit. on pp. 1, 5, 9, 10, 18, 29).

[Sra+16] S. Sra et al. “Adadelay: delay adaptive distributed stochastic optimization”. In: *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*. 16. Cadiz, Spain, 2016, pp. 957–965 (cit. on pp. 2, 14).

[Ste81] C. M. Stein. “Estimation of the mean of a multivariate normal distribution”. In: *The Annals of Statistics* 9.6 (Nov. 1981), pp. 1135–1151 (cit. on pp. 127, 131).

[TDT15a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “A perturbed linear mixing model accounting for spectral variability”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Nice, France, Sept. 2015, pp. 819–823.

[TDT15b] 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*. in French. Lyon, France, Sept. 2015.

[TDT16a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Hyperspectral unmixing with spectral variability using a perturbed linear mixing model”. In: *IEEE Trans. Signal Process.* 64.2 (Jan. 2016), pp. 525–538 (cit. on p. 17).

[TDT16b] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Online unmixing of multitemporal hyperspectral images accounting for spectral variability”. In: *IEEE Trans. Image Process.* 25.9 (Sept. 2016), pp. 3979–3990 (cit. on p. 75).

[TDT16c] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Unmixing multitemporal hyperspectral images with variability: an online algorithm”. In: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. Shanghai, China, Mar. 2016, pp. 3351–3355.

[TDT17a] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “A hierarchical Bayesian model accounting for endmember variability and abrupt spectral changes to unmix multitemporal hyperspectral images”. In: (May 2017). submitted. URL: `http://thouvenin.perso.enseeiht.fr/papers/Thouvenin2017.pdf` (cit. on p. 43).

[TDT17b] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Partially asynchronous distributed unmixing of hyperspectral images”. In: (2017). URL: `https://arxiv.org/abs/1710.02574` (cit. on p. 99).

[TDT17c] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Une approche distribuée asynchrone pour la factorisation en matrices non-négatives – application au démélange hyperspectral”. In: *Actes du XXVIième Colloque GRETSI*. in French. Juan-les-Pins, France, Sept. 2017.

[TDT17d] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. “Unmixing Multitemporal Hyperspectral Images Accounting for Smooth and Abrupt Variations”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. Kos, Greece, Sept. 2017, pp. 2442–2446.

[Tib96] R. Tibshirani. “Regression shrinkage and selection via the LASSO”. In: *J. Roy. Stat. Soc. Ser. B* 58.1 (1996), pp. 267–288 (cit. on p. 49).

[TRB17] C. G. Tsinos, A. A. Rontogiannis, and K. Berberidis. “Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization”. In: *IEEE Trans. Comput. Imag.* 3.2 (Apr. 2017), pp. 160–174 (cit. on pp. 110, 111, 118, 120, 121).

[Uez+16a] T. Uezato et al. “A novel endmember bundle extraction and clustering approach for capturing spectral variability within endmember classes”. In: *IEEE Trans. Geosci. Remote Sens.* 54.11 (Nov. 2016), pp. 6712–6731 (cit. on p. 10).

[Uez+16b] T. Uezato et al. “A novel spectral unmixing method incorporating spectral variability within endmember classes”. In: *IEEE Trans. Geosci. Remote Sens.* 54.5 (May 2016), pp. 2812–2831 (cit. on p. 10).

[Uez+16c] T. Uezato et al. “Incorporating spatial information and endmember variability into unmixing analyses to improve abundance estimates”. In: *IEEE Trans. Image Process.* 25.12 (Dec. 2016), pp. 5563–5575 (cit. on p. 10).

[Veg+14] M. A. Veganzones et al. “A new extended linear mixing model to address spectral variability”. In: *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*. Lausanne, Switzerland, June 2014 (cit. on pp. 1, 5, 10, 17, 18).

[Veg+15a] M. A. Veganzones et al. “Multilinear spectral unmixing of hyperspectral multiangle images”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. 23. Nice, France, Aug. 2015, pp. 744–48 (cit. on p. 11).

[Veg+15b] M. A. Veganzones et al. “Nonnegative tensor cp decomposition of hyperspectral data”. In: *IEEE Trans. Geosci. Remote Sens.* 54.5 (May 2015), pp. 2577–2588 (cit. on p. 11).

[VS13] J. P. Vila and P. Schniter. “Expectation-Maximization Gaussian-mixture approximate message passing”. In: *IEEE Trans. Signal Process.* 61.19 (Oct. 2013), pp. 4658–4672 (cit. on p. 49).

[Wei+16] Q. Wei et al. “Multiband image fusion based on spectral unmixing”. In: *IEEE Trans. Geosci. Remote Sens.* 54.12 (Dec. 2016), pp. 7236–7249 (cit. on pp. 128, 132).

[Wei15] Q. Wei. “Bayesian fusion of multi-band images: a powerful tool for super-resolution”. PhD thesis. Institut National Polytechnique de Toulouse, 2015. URL: http://www.theses.fr/2015INPT0059 (cit. on pp. 1, 5).

[Wri15] S. J. Wright. “Coordinate descent algorithms”. In: *Math. Program., Ser. B* 151.1 (June 2015), pp. 3–34 (cit. on p. 127).

[WYZ15] Y. Wang, W. Yin, and J. Zeng. “Global convergence of admm in nonconvex nonsmooth optimization”. In: *UCLA CAM Report 15-62* (Nov. 2015). URL: https://arxiv.org/abs/1511.06324 (cit. on p. 101).

[Yan+16] Y. Yang et al. “A parallel decomposition method for nonconvex stochastic multi-agent optimization problems”. In: *IEEE Trans. Signal Process.* 64.11 (June 2016), pp. 2949–2964 (cit. on pp. 99, 100).

[Yu13] Y.-L. Yu. “On decomposing the proximal map”. In: *Adv. in Neural Information Processing Systems*. 2013, pp. 91–99. URL: http://papers.nips.cc/paper/4863-on-decomposing-the-proximal-map.pdf (cit. on pp. 127, 131, 142).

[YZP17] N. Yokoya, X. X. Zhu, and A. Plaza. “Multisensor Coupled Spectral Unmixing for Time-Series Analysis”. In: *IEEE Trans. Geosci. Remote Sens.* 55.5 (May 2017), pp. 2842–2857 (cit. on pp. 11, 44, 45, 76, 108, 109, 116, 128, 132).

[ZH14] A. Zare and K. C. Ho. “Endmember variability in hyperspectral imagery”. In: *IEEE Signal Process. Mag.* 31.1 (Jan. 2014), pp. 95–104 (cit. on pp. 1, 5, 7, 8, 10, 19).