Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing

Clémence Prévost, Ricardo A Borsoi, Konstantin Usevich, David Brie, José C. M. Bermudez, Cédric Richard

Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing

Clémence Prévost, *Student Member, IEEE*, Ricardo A. Borsoi, *Student Member, IEEE*, Konstantin Usevich, *Member, IEEE*, David Brie, *Member, IEEE*, José C. M. Bermudez, *Senior Member, IEEE*, Cédric Richard, *Senior Member, IEEE*

***Abstract*—In this paper, we propose to jointly solve the hyperspectral super-resolution and hyperspectral unmixing problems using a coupled LL1 block-tensor decomposition. We focus on the specific case of spectral variability occurring between the observed low-resolution images. Exact recovery conditions are provided. We propose two algorithms: an unconstrained one and another one subject to non-negativity constraints, to solve the problems at hand. We showcase performance of the proposed approach on a set of synthetic and semi-real images.**

***Index Terms*—Hyperspectral super-resolution, spectral variability, hyperspectral unmixing, image fusion, tensor decompositions.**

I. Introduction

A. Background

Hyperspectral devices are able to sample the electromagnetic spectrum for hundred of wavelengths, allowing for the acquisition of hyperspectral images (HSIs) that possess high spectral resolution. The composition of each pixel in an HSI can frequently be approximated by a linear mixture of a small number of spectral signatures known as endmembers. This representation is known as the linear mixing model (LMM), and allows for identification of the materials in a scene, a process termed unmixing. Many hyperspectral unmixing (HU) approaches have been proposed (see [1], [2], [3] and references therein). However, the natural tradeoff between spatial and spectral resolution forces the HSIs to have few number of pixels [4]. On the other hand, multispectral devices produce multispectral images (MSIs) with high spatial resolution, at the cost of the spectral resolution; indeed, MSIs only include a restricted number of spectral bands.

Recently, the hyperspectral super-resolution (HSR) problem [5] was formulated to circumvent the physical limitations of each device. This problem aims at recovering a super-resolution image (SRI) that possesses both high spatial and high spectral resolutions from co-registered HSI and MSI of the same scene, without the need for any higher-resolution sensor.

Many approaches have been proposed to solve the HSR problem. Early matrix-based approaches include coupled non-negative matrix factorization (CNMF) [6], methods based on solving Sylvester equations [7], Bayesian approaches (HySure) [8], FUMI [9], to name a few. Motivated by the LMM, most of these methods are based on a coupled low-rank factorization of the matricized HSI and MSI. One advantage of using the LMM is that the latent factors of a properly chosen decomposition of the recovered SRI can be seen as the spectral signatures and abundance maps corresponding to the underlying materials in the image, provided that they are entry-wise non-negative. For instance, in [10], the HSR problem is solved by recovering high-resolution spectra from the HSI, and the abundance maps are estimated jointly from the two low-resolution images. In [6], high-resolution abundance matrices are recovered from the MSI while high-resolution spectral signatures are obtained from the HSI. Thus some matrix approaches are suitable for the HSR-HU problem, which consists of recovering the underlying SRI by means of a physically-informed low-rank approximation. However, for these methods, identifiability of the mixing model can only be obtained under additional constraints on the low-rank factors [11], [12]. Regarding the fusion problem, some matrix-based approaches provide exact recovery conditions for the SRI in noiseless cases, see for example [10] that promotes sparsity of the low-rank factors. However, these recovery guarantees are usually conditioned to the incorporation of priors on the low-rank factors as well. In the absence of such hypotheses, a bound on the recovery error can be obtained [13]. As a result, to the best of our understanding, recovery conditions for the joint HSR-HU problem in the literature only consider specific classes of problems. Furthermore, related algorithms often suffer from high computational complexity.

More recently, tensor-based approaches were proposed for solving the HSR problem, exploiting the inherent 3-dimensional nature of HSIs and MSIs. The works of [14], [15] formulate the HSR problem as a coupled canonical polyadic (CP) decomposition (CPD), while a coupled multilin-

This work was partly supported by the ANR (Agence Nationale de Recherche) under grant LeaFleT (ANR-19-CE23-0021), by the National Council for Scientific and Technological Development (CNPq) under grants 304250/2017-1, 409044/2018-0, 141271/2017-5 and 204991/2018-8. The work of C. Richard was supported through the UCA JEDI Investments in the Future project with the reference number ANR-15-IDEX-0001.

We also thank the GdR CNRS ISIS for supporting the collaboration between R.A. Borsoi and C. Prévost, K. Usevich and D. Brie through a *bourse de mobilité*.

C. Prévost, K. Usevich and D. Brie are with Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy, France. e-mail: firstname.lastname@univ-lorraine.fr.

R.A. Borsoi is with the Department of Electrical Engineering, Federal University of Santa Catarina (DEE–UFSC), Florianópolis, SC, Brazil, and with the Lagrange Laboratory, Université Côte d'Azur, Nice, France. e-mail: raborsoi@gmail.com.

J.C.M. Bermudez is with the DEE–UFSC, Florianópolis, SC, Brazil, and with the Graduate Program on Electronic Engineering and Computing, Catholic University of Pelotas (UCPel) Pelotas, Brazil. e-mail: j.bermudez@ieee.org.

C. Richard is with the Université Côte d'Azur, Nice, France (e-mail: cedric.richard@unice.fr), Lagrange Laboratory (CNRS, OCA).

ear Tucker decomposition is used in [16]. These formulations proved state-of-the-art practical performance as well as exact noiseless recovery guarantees for the SRI, often at a low computational cost. However, the latent factors of the CP or Tucker decompositions lack physical interpretation, and thus the aforementioned methods cannot be used for hyperspectral unmixing. Motivated by the usefulness of tensor methods, an approach based on block-tensor decomposition (BTD) was recently proposed [17], [18]; the main advantage of this approach is to link the latent factors of the BTD model to abundance matrices and spectral signatures used in the HU task, under additional non-negativity priors on the low-rank factors.

Most of these approaches however share a common limitation: they assume that the acquisition conditions of the HSI and MSI are the same. In practice, there exist very few optical satellites that carry both HS and MS sensors [19], [20]. Thus, combining an HSI and MSI acquired onboard of different missions has become a task of prime interest [21], [22]. Since the HSI and MSI are acquired at different time instants, they can differ by, e.g., illumination, atmospheric or seasonal changes [23], which can cause variations in the underlying spectra of the acquired images, and impact negatively the fusion and unmixing algorithms. In [24], a super-resolution method was proposed, combining the HSI and MSI accounting for seasonal spectral variability. Using a low-rank matrix formulation, the spectral signatures underlying the HSI and MSI are allowed to be different from each other, with variations introduced by a set of multiplicative scaling factors [25]. This algorithm led to significant performance improvements when the HSI and MSI are subject to spatially uniform seasonal or acquisition variations. However, the algorithm in [24] presents high computation times and does not offer any theoretical guarantees. In [26], two tensor-based algorithms based on Tucker and block-term decompositions were proposed, accounting for inter-image variations between the HSI and MSI. The approaches of [26] proved state-of-the-art performance for the HSR task, as well as guarantees for noiseless uniqueness of the target images. Unfortunately, the decomposition factors were not suitable for unmixing.

In this paper, we propose to formulate the HSR problem as a coupled BTD of the HSI and MSI, accounting for spectral variability between the endmembers. Inspired by the work of [17], [18], we propose some guarantees for noiseless unique recovery of the SRI and its latent factors based on the LMM. In particular, our noiseless recovery conditions do not required additional constraints on the low-rank factors, contrary to matrix-based models. We also propose a procedure that aims at recovering the SRI and estimating its latent factors. Our experiments on a set of synthetic and semi-real datasets prove competitive performance of the proposed approach for the HSR and HU tasks, compared to traditional unmixing algorithms.

This paper is organized as follows. The remainder of Section I is devoted to tensor algebra preliminaries. Section II introduces the tensor degradation model, as well as the coupled low-rank model accounting for spectral variability. Section III addresses recoverability analysis for the joint HSR-HU task. Section IV describes the proposed algorithms and their computational complexity. Finally, Sections V and VI contain numerical experiments for the HSR and HU tasks, respectively.

B. Definitions and notations

In this paper, we mainly follow the notations of [27], [28]. We use the following fonts: lower (a) or uppercase (A) plain font for scalars, boldface lowercase ($\boldsymbol{a}$) for vectors, boldface uppercase ($\boldsymbol{A}$) for matrices and calligraphic ($\boldsymbol{\mathcal{A}}$) for tensors. The elements of vectors, matrices and tensors are denoted as a_i, $A_{i,j}$ and $\mathcal{A}_{i_1,\ldots,i_N}$, respectively. For a matrix $\boldsymbol{A}$, its transpose is denoted by $\boldsymbol{A}^\mathsf{T}$. We use the notation $\boldsymbol{I}_N$ for the $N \times N$ identity matrix and $\boldsymbol{0}_{L\times K}$ for the $L \times K$ matrix of zeros. The notation $\mathbb{1}_L$ denotes an all-ones vector of size $L \times 1$. The symbol $\boxtimes$ and $\odot$ denote the Kronecker and Khatri-Rao products, respectively. We denote the Hadamard (element-wise) product by $\boxdot$. We use $\mathrm{vec}\{\cdot\}$ for the standard column-major vectorization of a matrix or a tensor.

Each dimension of a tensor is called a mode. A mode-p fiber of tensor $\boldsymbol{\mathcal{X}}$ is a vector of $\boldsymbol{\mathcal{X}}$ obtained by fixing all but the p-th dimension. Similarly, a slab or slice of a tensor $\boldsymbol{\mathcal{X}}$ is a matrix whose columns are the vectors of $\boldsymbol{\mathcal{X}}$ obtained by fixing all but two of its modes. In this paper, we restrict to the scope of three-dimensional tensors.

Definition I.1. ***Outer product*** **–** *The outer product between three vectors $\boldsymbol{a} \in \mathbb{R}^I$, $\boldsymbol{b} \in \mathbb{R}^J$, $\boldsymbol{c} \in \mathbb{R}^K$ is an order-3 tensor of rank 1 defined as $\boldsymbol{\mathcal{X}} = \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \in \mathbb{R}^{I\times J\times K}$. Each element of $\boldsymbol{\mathcal{X}}$ is accessed as $\mathcal{X}_{i,j,k} = a_i b_j c_k$.*

Definition I.2. ***Tensor unfoldings*** **–** *The mode-p unfolding of a tensor $\boldsymbol{\mathcal{X}}$, denoted by $\boldsymbol{X}^{(p)}$, is the matrix whose rows are the p-mode fibers of $\boldsymbol{\mathcal{X}}$, ordered according to the vectorization order. For a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, we have $\boldsymbol{X}^{(1)} \in \mathbb{R}^{JK\times I}$, $\boldsymbol{X}^{(2)} \in \mathbb{R}^{IK\times J}$ and $\boldsymbol{X}^{(3)} \in \mathbb{R}^{IJ\times K}$.*

Definition I.3. ***Mode product*** **–** *The mode-p product between a tensor $\boldsymbol{\mathcal{X}}$ and a matrix $\boldsymbol{M}$ is denoted by $\boldsymbol{\mathcal{X}} \bullet_p \boldsymbol{M}$ and is evaluated such that each mode-p fiber of $\boldsymbol{\mathcal{X}}$ is multiplied by $\boldsymbol{M}$. For instance, the elements of the mode-1 product between $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ and $\boldsymbol{M} \in \mathbb{R}^{L\times I}$ are accessed as $(\boldsymbol{\mathcal{X}} \bullet_1 \boldsymbol{M})_{\ell,j,k} = \sum_i \mathcal{X}_{i,j,k} \boldsymbol{M}_{\ell,i}$, $\ell \in \{1,\ldots,L\}$. Moreover, it holds that $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \bullet_k \boldsymbol{M} \Leftrightarrow \boldsymbol{Y}^{(k)} = \boldsymbol{X}^{(k)} \boldsymbol{M}^\mathsf{T}$.*

C. Block-term decomposition with ranks $(L, L, 1)$

In this subsection, we introduce the block-term decomposition with ranks $(L, L, 1)$, that we will use to build our model. We also recall some sufficient uniqueness conditions for this decomposition, as well as useful properties.

Definition I.4. ***Block-term decomposition*** **–** *An order-3 tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ generally admits a block-term decomposition (BTD) with ranks $(L, L, 1)$ (LL1-BTD) as*

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left(\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}\right) \otimes \boldsymbol{c}_r, \tag{1}$$

where $\boldsymbol{A}_r \in \mathbb{R}^{I\times L}$, $\boldsymbol{B}_r \in \mathbb{R}^{J\times L}$, and $\boldsymbol{c}_r \in \mathbb{R}^K$, for $r \in \{1,\ldots,R\}$. Moreover, we denote $\boldsymbol{A} = [\boldsymbol{A}_1,\ldots,\boldsymbol{A}_R] \in \mathbb{R}^{I\times LR}$, $\boldsymbol{B} = [\boldsymbol{B}_1,\ldots,\boldsymbol{B}_R] \in \mathbb{R}^{J\times LR}$ and $\boldsymbol{C} = [\boldsymbol{c}_1,\ldots,\boldsymbol{c}_R] \in \mathbb{R}^{K\times R}$.

Theorem I.5. *[29, Theorem 4.7] Let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ denote an LL1-BTD of a tensor $\boldsymbol{\mathcal{X}}$ for $r \in \{1,\ldots,R\}$ as in (1). As-*

sume that $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ *are drawn from certain joint absolutely continuous distributions. If* $IJ \geq L^2R$ *and*

$$\min\left(\lfloor\frac{I}{L}\rfloor, R\right) + \min\left(\lfloor\frac{J}{L}\rfloor, R\right) + \min(K, R) \geq 2R + 2,$$

then $\boldsymbol{A}_r\boldsymbol{B}_r^{\mathsf{T}}$ *and* $\boldsymbol{c}_r$ *are essentially unique almost surely for* $r \in \{1, \ldots, R\}$.

Definition I.6. ***Partition-wise Khatri-Rao product*** **–** *The partition-wise Khatri-Rao product between two partitioned matrices* $\boldsymbol{A}$ *and* $\boldsymbol{C}$ *defined as above can be expressed as*

$$\boldsymbol{C} \odot_p \boldsymbol{A} = [\boldsymbol{c}_1 \boxtimes \boldsymbol{A}_1, \ldots, \boldsymbol{c}_R \boxtimes \boldsymbol{A}_R] \in \mathbb{R}^{IK \times LR}.$$

Property 1. ***Tensor unfoldings and LL1*** **–** *Using the above notations, the unfoldings of a tensor* $\boldsymbol{\mathcal{X}}$ *admitting an LL1-BTD as above can be expressed as*

$$\begin{aligned}
\boldsymbol{X}^{(1)} &= (\boldsymbol{C} \odot_p \boldsymbol{B})\, \boldsymbol{A}^{\mathsf{T}}, \\
\boldsymbol{X}^{(2)} &= (\boldsymbol{C} \odot_p \boldsymbol{A})\, \boldsymbol{B}^{\mathsf{T}}, \\
\boldsymbol{X}^{(3)} &= [(\boldsymbol{A}_1 \odot \boldsymbol{B}_1)\, \mathbb{1}_L, \ldots, (\boldsymbol{A}_R \odot \boldsymbol{B}_R)\, \mathbb{1}_L]\, \boldsymbol{C}^{\mathsf{T}}.
\end{aligned}$$

II. Proposed model

A. Degradation model and indeterminacies

We consider an HSI data cube $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H \times J_H \times K}$ and an MSI data cube $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I \times J \times K_M}$. While the scalars K and K_M denote the spectral dimensions, (I, J) (resp. (I_H, J_H)) stand for the spatial dimensions of the data cubes. We suppose that the spatial resolution of the MSI is higher than that of the HSI (i.e., $I_H < I$ and $J_H < J$), while its spectral resolution is lower ($K_M < K$). Most previous works [14]–[18] considered the low resolution images as degraded versions of a single SRI $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I \times J \times K}$, that possesses high spatial and spectral resolutions. This model can be expressed as mode product of $\boldsymbol{\mathcal{Z}}$ with degradation matrices such that:

$$\begin{cases}
\boldsymbol{\mathcal{Y}}_H &= \boldsymbol{\mathcal{Z}} \bullet_1 \boldsymbol{P}_1 \bullet_2 \boldsymbol{P}_2 + \boldsymbol{\mathcal{E}}_H, \\
\boldsymbol{\mathcal{Y}}_M &= \boldsymbol{\mathcal{Z}} \bullet_3 \boldsymbol{P}_3 + \boldsymbol{\mathcal{E}}_M,
\end{cases} \tag{2}$$

where the tensors $\boldsymbol{\mathcal{E}}_H$ and $\boldsymbol{\mathcal{E}}_M$ are additive noise terms. The matrix $\boldsymbol{P}_3 \in \mathbb{R}^{K_M \times K}$ contains the spectral response functions (SRF) for each band of the MSI sensor. The spatial degradation matrices $\boldsymbol{P}_1 \in \mathbb{R}^{I_H \times I}$ and $\boldsymbol{P}_2 \in \mathbb{R}^{J_H \times J}$ perform Gaussian blurring and downsampling along each spatial dimension, *i.e.* we suppose that the spatial degradation operation is separable, as in the commonly used Wald's protocol [30].

However, this model implies that the acquisition conditions of $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$ are the same, and thus ignores any variability phenomenon. Variations in atmospheric, seasonal or illumination conditions [31], [32] between the HSI and MSI motivate the need for more flexible models.

As a result, in this paper, we adopt a more general approach. As in [26], we consider two different SRIs $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I \times J \times K}$ and $\widetilde{\boldsymbol{\mathcal{Z}}} \in \mathbb{R}^{I \times J \times K}$, underlying the HSI and MSI, respectively. The SRIs $\boldsymbol{\mathcal{Z}}$ and $\widetilde{\boldsymbol{\mathcal{Z}}}$ contain possibly different spectral signatures and can be linked as

$$\widetilde{\boldsymbol{\mathcal{Z}}} = \boldsymbol{\mathcal{Z}} + \boldsymbol{\Psi}, \tag{3}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{I \times J \times K}$ is a tensor of variability. This leads to the following extension of model (2):

$$\begin{cases}
\boldsymbol{\mathcal{Y}}_H &= \boldsymbol{\mathcal{Z}} \bullet_1 \boldsymbol{P}_1 \bullet_2 \boldsymbol{P}_2 + \boldsymbol{\mathcal{E}}_H, \\
\boldsymbol{\mathcal{Y}}_M &= \widetilde{\boldsymbol{\mathcal{Z}}} \bullet_3 \boldsymbol{P}_3 + \boldsymbol{\mathcal{E}}_M.
\end{cases} \tag{4}$$

In this framework, the HSR problem consists in the following:

$$\begin{cases}
\text{find } \boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I \times J \times K} \text{ and } \boldsymbol{\Psi} \in \mathbb{R}^{I \times J \times K}, \\
\text{such that (3) and (4) are satisfied.}
\end{cases} \tag{5}$$

However, this problem is severely ill-posed. Indeed, the presence of the variability tensor $\boldsymbol{\Psi}$ makes problem (5) ambiguous [26], as one cannot separate easily $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\Psi}$ from $\widetilde{\boldsymbol{\mathcal{Z}}}$. We recall the following theorem [26]:

Theorem II.1. *[26, Theorem 1.a)] Suppose that the HSI and MSI are generated according to eq. (4) and that the observation noise is zero (i.e.* $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \boldsymbol{0}$*). If either* $\boldsymbol{P}_1$*,* $\boldsymbol{P}_2$ *or* $\boldsymbol{P}_3$ *have non-trivial nullspace,, then* $(\boldsymbol{\mathcal{Z}}, \boldsymbol{\Psi})$ *cannot be uniquely recovered from* $\boldsymbol{\mathcal{Y}}_H$ *and* $\boldsymbol{\mathcal{Y}}_M$.

Thus, without any prior information to the structure of $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\Psi}$, the target images cannot be recovered uniquely. In what follows, we introduce a low-rank tensor model to circumvent the fundamental ambiguities of the proposed degradation model. Indeed, while Theorem II.1 indicates conditions under which $(\boldsymbol{\mathcal{Z}}, \boldsymbol{\Psi})$ cannot be recovered uniquely, a wisely chosen low-rank decomposition might allow for unique recovery of portions of the tensors. In [26], approaches based on model (4) were proposed, using a coupled Tucker approximation. However, the latent multilinear factors were not interpretable as a mixing model. In this paper we propose to use LL1-BTD, whose factors are suitable for physical interpretation.

Different from [26] (where both spatial and spectral variabilities were considered), we consider that variability only impacts the spectral dimensions of the SRI. This assumption is reasonable, since spectral variability can occur even with short acquisition time differences. Addressing spectral variability only allows to visualize which wavelengths are more impacted by different acquisition conditions for each specific material. This hypothesis also allows for design of simple algorithms, with sometimes less restrictive noiseless recoverability guarantees than the ones from [26][1]. As a result, the proposed model is more suitable for scenarios with low spatial variability. We will show in Sections V and VI that the proposed model is able to address large acquisition time differences as well.

B. LL1-BTD mixing model for the underlying SRIs

Each pixel (or mode-3 fiber) of $\boldsymbol{\mathcal{Z}}$ (and therefore, of $\boldsymbol{\mathcal{Y}}_H$) can be represented as a linear mixture of a small number R of pure spectral signatures [33]. This property can be incorporated in a physically-informed model, allowing to perform both image fusion and unmixing at the same time. Thus, as in [17], we can model the third-mode unfolding of $\boldsymbol{\mathcal{Z}}$ using the LMM as:

$$\boldsymbol{Z}^{(3)} = \boldsymbol{S}\boldsymbol{C}^{\mathsf{T}} \in \mathbb{R}^{IJ \times K}, \tag{6}$$

where $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R] \in \mathbb{R}^{K \times R}$ is a matrix containing the spectral signatures of the R endmembers underlying the SRI. The matrix $\boldsymbol{S} = [\mathrm{vec}\{\boldsymbol{S}_1\}, \ldots, \mathrm{vec}\{\boldsymbol{S}_R\}] \in \mathbb{R}^{IJ \times R}$ contains the vectorized abundance maps of each material. In the LMM, $\boldsymbol{C}$ and $\boldsymbol{S}$ are assumed to be entry-wise non-negative.

In some traditional unmixing methods (see e.g. [3]), the sum-to-one constrained abundance matrices is also enforced.

[1] In [26], recovery of the full variability tensor using a Tucker approximation often results in restrictive multilinear ranks.

However, spatial illumination changes frequently introduce scaling variations in each pixel. Moreover, non-negativity constraints can be transformed equivalently to generalized sum-to-one constraint, as specified in [34]. As a result, we do not consider this additional constraint in this work.

We assume as in [17] that the abundance maps admit low rank L. The hypothesis of low-rank abundance matrices is reasonable, since the two spatial dimensions are often correlated along the rows and columns, respectively. Thus we have

$$\boldsymbol{S}_r \approx \boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T} \in \mathbb{R}^{I \times J},$$

where $\boldsymbol{A}_r \in \mathbb{R}^{I \times L}$ and $\boldsymbol{B}_r \in \mathbb{R}^{J \times L}$ admit rank L.

Reshaping $\boldsymbol{Z}^{(3)}$ into tensor format yields the following:

$$\boldsymbol{\mathcal{Z}} = \sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}) \otimes \boldsymbol{c}_r. \tag{7}$$

The above model can be seen as an LL1-BTD of the tensor $\boldsymbol{\mathcal{Z}}$ with factors $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_r]$, $\boldsymbol{B} = [\boldsymbol{B}_1, \ldots, \boldsymbol{B}_r]$ and $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_r]$. Traditional unmixing algorithms aim at recovering $\{\boldsymbol{S}_r = \boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}\}_{r=1}^R$ and $\boldsymbol{C}$ from the mixed pixels in $\boldsymbol{\mathcal{Z}}$. Here, since only $\boldsymbol{\mathcal{Y}}_H$ is observed with high spectral resolution, traditional unmixing is only able to recover spatially degraded versions of the abundance maps [18]

$$\boldsymbol{P}_1 \boldsymbol{S}_r \boldsymbol{P}_2^\mathsf{T} \in \mathbb{R}^{I_H \times J_H} \text{ for } r \in \{1, \ldots, R\}. \tag{8}$$

Differently from those works, it is possible to seek for abundance maps at a higher spatial resolution by exploiting the high spatial resolution of the MSI. However, the spectral variability between the HSI and MSI has to be modeled first.

C. Modeling spectral variability

In [25], the generalized linear mixing model (GLMM) was proposed to model the spectra underlying the MSI as

$$\widetilde{\boldsymbol{C}} = \boldsymbol{\psi}_{\text{multi}} \boxdot \boldsymbol{C}, \tag{9}$$

where $\boldsymbol{\psi}_{\text{multi}} \in \mathbb{R}^{K \times R}$ is a matrix of positive scaling factors. In this paper, motivated by model (2), we propose to use an equivalent additive model due to its better mathematical tractability:

$$\widetilde{\boldsymbol{C}} = \boldsymbol{\psi} + \boldsymbol{C}, \tag{10}$$

where $\boldsymbol{\psi} \in \mathbb{R}^{K \times R}$ is different from $\boldsymbol{\psi}_{\text{multi}}$.

Since we allow only spectral variability to be present, the variability tensor $\boldsymbol{\Psi}$ also admits an LL1-BTD with the same factors $\boldsymbol{A}$ and $\boldsymbol{B}$ as for the SRI $\boldsymbol{\mathcal{Z}}$, but with spectral factor $\boldsymbol{\psi} \in \mathbb{R}^{K \times R}$ representing spectral variability. This allows to write $\widetilde{\boldsymbol{\mathcal{Z}}}$ as

$$\widetilde{\boldsymbol{\mathcal{Z}}} = \underbrace{\sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}) \otimes \boldsymbol{c}_r}_{\boldsymbol{\mathcal{Z}}} + \underbrace{\sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}) \otimes \boldsymbol{\psi}_r}_{\boldsymbol{\Psi}}$$

$$\widetilde{\boldsymbol{\mathcal{Z}}} = \sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}) \otimes \widetilde{\boldsymbol{c}}_r, \tag{11}$$

where $\widetilde{\boldsymbol{c}}_r = \boldsymbol{\psi}_r + \boldsymbol{c}_r$ is the r-th column of $\widetilde{\boldsymbol{C}}$.

From the above formulations, we can finally express eq. (4) as a coupled LL1-BTD:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H &= \sum_{r=1}^{R} (\boldsymbol{P}_1 \boldsymbol{A}_r (\boldsymbol{P}_2 \boldsymbol{B}_r)^\mathsf{T}) \otimes \boldsymbol{c}_r + \boldsymbol{\mathcal{E}}_H, \\ \boldsymbol{\mathcal{Y}}_M &= \sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}) \otimes \boldsymbol{P}_3 \widetilde{\boldsymbol{c}}_r + \boldsymbol{\mathcal{E}}_M. \end{cases} \tag{12}$$

Thus we define the joint HSR-HU problem as:

$$\begin{cases} \text{find } \{\boldsymbol{A}_r, \boldsymbol{B}_r\}_{r=1}^R, \boldsymbol{C}, \widetilde{\boldsymbol{C}}, \text{ such that (12) is satisfied,} \\ \text{s. to } \{\boldsymbol{A}_r \boldsymbol{B}_r^\mathsf{T}\}_{r=1}^R \geq \boldsymbol{0}, \boldsymbol{C} \geq \boldsymbol{0}, \widetilde{\boldsymbol{C}} \geq \boldsymbol{0}, \end{cases} \tag{13}$$

where for a matrix $\boldsymbol{X}$, the notation $\boldsymbol{X} \geq \boldsymbol{0}$ means that $\boldsymbol{X}$ is entry-wise non-negative.

III. Recoverability analysis

In fact, the HSR-HU problem (13) aims at an SRI $\boldsymbol{\mathcal{Z}}$ and a variability tensor $\boldsymbol{\Psi}$ underlying the HSI in MSI, admitting a coupled LL1-BTD as in (12). Regarding the unmixing task, the latent factors underlying the images must also be recovered uniquely. In this section, we provide a noiseless recoverability analysis for the SRI $\boldsymbol{\mathcal{Z}}$ and variability tensor $\boldsymbol{\Psi}$. We show that our results hold for both image recovery and estimation of the latent factors based on the LMM.

Theorem III.1. *Assume that the SRIs $\boldsymbol{\mathcal{Z}}$ and $\widetilde{\boldsymbol{\mathcal{Z}}}$ admit BTDs as in (7) (resp. (11)), that the HSI and MSI follow the coupled model (12), and that $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \boldsymbol{0}$. Suppose that $\{\boldsymbol{A}_r, \boldsymbol{B}_r\}_{r=1}^R, \boldsymbol{C}, \widetilde{\boldsymbol{C}}$ are drawn from any absolutely continuous joint distributions and that $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ are full row rank. Let $\{\boldsymbol{A}_r^*, \boldsymbol{B}_r^*\}_{r=1}^R, \boldsymbol{C}^*, \widetilde{\boldsymbol{C}}^*$ denote any solution to problem (13). Then with probability one, the true SRI $\boldsymbol{\mathcal{Z}}$ and degraded SRI $\widetilde{\boldsymbol{\mathcal{Z}}} \bullet_3 \boldsymbol{P}_3$ are uniquely recovered by*

$$\boldsymbol{\mathcal{Z}} = \sum_{r=1}^{R} (\boldsymbol{A}_r^* (\boldsymbol{B}_r^*)^\mathsf{T}) \otimes \boldsymbol{c}_r^*,$$

$$\widetilde{\boldsymbol{\mathcal{Z}}} \underset{3}{\bullet} \boldsymbol{P}_3 = \sum_{r=1}^{R} (\boldsymbol{A}_r^* (\boldsymbol{B}_r^*)^\mathsf{T}) \otimes \boldsymbol{P}_3 \widetilde{\boldsymbol{c}}_r^*,$$

if $I_H J_H \geq LR$, $IJ \geq L^2 R$ and

$$\min\left(\lfloor \frac{I}{L} \rfloor, R\right) + \min\left(\lfloor \frac{J}{L} \rfloor, R\right) + \min(K_M, R) \geq 2R + 2.$$

Let us first recall the following lemma:

Lemma III.2. *[14, Lemma 1] Let us denote $\widetilde{\boldsymbol{A}} = \boldsymbol{P}\boldsymbol{A} \in \mathbb{R}^{I' \times L}$, where $\boldsymbol{P} \in \mathbb{R}^{I' \times I}$ is full row rank and $\boldsymbol{A} \in \mathbb{R}^{I \times L}$ is drawn from any absolutely continuous joint distribution. Then $\widetilde{\boldsymbol{A}}$ follows an absolutely continuous joint distribution.*

We can now derive the proof for Theorem III.1.

Proof. Let $\{\boldsymbol{A}_r, \boldsymbol{B}_r\}_{r=1}^R, \boldsymbol{C}, \widetilde{\boldsymbol{C}}$ denote the groundtruth factors of the SRI tensors and let $\{\boldsymbol{A}_r^*, \boldsymbol{B}_r^*\}_{r=1}^R, \boldsymbol{C}^*, \widetilde{\boldsymbol{C}}^*$ denote a solution to problem (13). Moreover, let $\{\boldsymbol{S}_r\}_{r=1}^R$ denote the

groundtruth abundance maps and $\{\boldsymbol{S}_r^*\}_{r=1}^R = \{\boldsymbol{A}_r^*(\boldsymbol{B}_r^*)^\mathsf{T}\}_{r=1}^R$. Then for $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \boldsymbol{0}$, it holds that

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_H &= \sum_{r=1}^R (\boldsymbol{P}_1\boldsymbol{A}_r(\boldsymbol{P}_2\boldsymbol{B}_r)^\mathsf{T}) \otimes \boldsymbol{c}_r \\ &= \sum_{r=1}^R (\boldsymbol{P}_1\boldsymbol{A}_r^*(\boldsymbol{P}_2\boldsymbol{B}_r^*)^\mathsf{T}) \otimes \boldsymbol{c}_r^*, \qquad (14)\end{aligned}$$

$$\boldsymbol{\mathcal{Y}}_M = \sum_{r=1}^R \boldsymbol{S}_r \otimes \boldsymbol{P}_3\widetilde{\boldsymbol{c}}_r = \sum_{r=1}^R \boldsymbol{S}_r^* \otimes \boldsymbol{P}_3\widetilde{\boldsymbol{c}}_r^*. \qquad (15)$$

Since by assumption, $\{\boldsymbol{A}_r, \boldsymbol{B}_r\}_{r=1}^R, \boldsymbol{C}, \widetilde{\boldsymbol{C}}$ are drawn from absolutely continuous joint distributions and $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ are full row rank, it follows from Lemma III.2 that $\{\boldsymbol{P}_1\boldsymbol{A}_r, \boldsymbol{P}_2\boldsymbol{B}_r\}_{r=1}^R, \boldsymbol{P}_3\widetilde{\boldsymbol{C}}$ follow certain absolutely continuous joint distributions.

Therefore, by Theorem I.5, the LL1-BTD of $\boldsymbol{\mathcal{Y}}_M$ is essentially unique almost surely if $IJ \geq L^2R$ and

$$\min\left(\lfloor\frac{I}{L}\rfloor, R\right) + \min\left(\lfloor\frac{J}{L}\rfloor, R\right) + \min(K_M, R) \geq 2R + 2.$$

This means that

$$\boldsymbol{S}^* = \boldsymbol{S}\boldsymbol{\Pi}\boldsymbol{\Lambda}, \qquad \boldsymbol{P}_3\widetilde{\boldsymbol{C}}^* = \boldsymbol{P}_3\widetilde{\boldsymbol{C}}\boldsymbol{\Pi}\boldsymbol{\Lambda}^{-1},$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Lambda}$ is a non-singular diagonal scaling matrix.

Next, let us define $\widetilde{\boldsymbol{S}} = (\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S}$. We can see that

$$\widetilde{\boldsymbol{S}}^* = \widetilde{\boldsymbol{S}}\boldsymbol{\Pi}\boldsymbol{\Lambda},$$

where $\widetilde{\boldsymbol{S}}^* = (\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S}^*$. From [35, Lemma 3.3] and the proof of [18, Theorem II], $\widetilde{\boldsymbol{S}}$ has full column rank almost surely if $I_H J_H \geq LR$.

Let us continue by considering $\boldsymbol{Y}_H^{(3)}$. From (14), we have

$$\boldsymbol{Y}_H^{(3)} = \widetilde{\boldsymbol{S}}\boldsymbol{C}^\mathsf{T} = \widetilde{\boldsymbol{S}}^*(\boldsymbol{C}^*)^\mathsf{T} = \widetilde{\boldsymbol{S}}\boldsymbol{\Pi}\boldsymbol{\Lambda}(\boldsymbol{C}^*)^\mathsf{T}.$$

Since $\widetilde{\boldsymbol{S}}$ has full column rank, we thus have $\boldsymbol{C}^* = \boldsymbol{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}^{-1}$.

Finally, we can express the third unfolding of the SRI $\boldsymbol{\mathcal{Z}}$ and degraded $\widetilde{\boldsymbol{\mathcal{Z}}} \bullet_3 \boldsymbol{P}_3$ as

$$\boldsymbol{Z}^{(3)} = \boldsymbol{S}^*(\boldsymbol{C}^*)^\mathsf{T}, \ (\widetilde{\boldsymbol{\mathcal{Z}}} \underset{3}{\bullet} \boldsymbol{P}_3)^{(3)} = \boldsymbol{S}^*(\boldsymbol{P}_3\widetilde{\boldsymbol{C}}^*)^\mathsf{T},$$

which completes the proof. □

Remark III.3. *In the proof for Theorem III.1, we can see that the low-rank factors $\boldsymbol{C}$ and $\boldsymbol{P}_3\widetilde{\boldsymbol{C}}$, as well as the vectorized abundance maps $\boldsymbol{S}$, can be uniquely identified up to permutation and scaling ambiguities. This means that our recoverability results also hold for the unmixing task of the HSR-HU problem (13), as it allows for noiseless unique recovery of the underlying abundance maps and spectra in $\boldsymbol{\mathcal{Z}}$ and $\widetilde{\boldsymbol{\mathcal{Z}}} \bullet_3 \boldsymbol{P}_3$. Thus Theorem III.1 proposed unique recovery conditions for the joint HSR-HU problem.*

Remark III.4. *We can see that knowledge of the spectral degradation matrix $\boldsymbol{P}_3$ is not needed[2] to establish uniqueness of the SRI $\boldsymbol{\mathcal{Z}}$. As a result, the proposed approach can be considered as blind in the spectral dimension. In fact, applying Theorem I.5 to the MSI $\boldsymbol{\mathcal{Y}}_M$ indicates that we can only recover $\boldsymbol{P}_3\widetilde{\boldsymbol{C}}$ uniquely, up to permutation and scaling ambiguities.*

Following (10), the variability matrix $\boldsymbol{\psi}$ can only be recovered from the MSI up to the spectral degradation $\boldsymbol{P}_3$ as

$$\boldsymbol{P}_3\boldsymbol{\psi} = \boldsymbol{P}_3(\widetilde{\boldsymbol{C}} - \boldsymbol{C}).$$

Thus the proposed model only allows to recover uniquely a spectrally-degraded version of the variability tensor, that is, $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$.

[2]Contrary to that of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$.

IV. Algorithms

In this section, we propose two spectrally-blind algorithms based on the LL1-BTD. The first one is unconstrained and solves the HSR problem only. The second one enforced non-negativity constraints on the factors of the mixing model and proposes a solution to the joint HSR-HU problem.

A. Unconstrained optimization

When only interested in the super-resolution problem, we only aim at recovering the SRI $\boldsymbol{\mathcal{Z}}$ and variability tensor $\boldsymbol{\Psi}$. In this framework, the latent LL1 factors do not need to be interpretable. Thus, we can consider unconstrained optimization. In the remaining of this paper, for simplicity, we denote $\widetilde{\boldsymbol{C}}_M = [\widetilde{\boldsymbol{c}}_{M,1}, \ldots, \widetilde{\boldsymbol{c}}_{M,R}] = \boldsymbol{P}_3\widetilde{\boldsymbol{C}}$.

As in [14], one possible approach for solving problem (5) is to formulate HSR as the following optimization problem:

$$\underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\widetilde{\boldsymbol{C}}_M}{\text{minimize}} \ \mathcal{J}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \widetilde{\boldsymbol{C}}_M), \qquad (16)$$

where

$$\begin{aligned}\mathcal{J}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \widetilde{\boldsymbol{C}}_M) &= \|\boldsymbol{\mathcal{Y}}_H - \sum_{r=1}^R (\boldsymbol{P}_1\boldsymbol{A}_r(\boldsymbol{P}_2\boldsymbol{B}_r)^\mathsf{T}) \otimes \boldsymbol{c}_r\|_F^2 \\ &+ \lambda\|\boldsymbol{\mathcal{Y}}_M - \sum_{r=1}^R (\boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T}) \otimes \widetilde{\boldsymbol{c}}_{M,r}\|_F^2,\end{aligned}$$

which is a non-convex cost function, and λ is a balance parameter that controls the respective weights on the HSI and MSI[3]. We circumvent this lack of convexity using a block coordinate descent (BCD) scheme: the latent factors are updated sequentially by solving unconstrained convex quadratic programs.

Below, we provide the general framework of the corresponding algorithm, denoted hereafter as BTD-Var.

The normalization step for the $\boldsymbol{C}$ and $\widetilde{\boldsymbol{C}}_M$ factors is meant to avoid underflow and overflow [36]. The updates for $\boldsymbol{A}$ and $\boldsymbol{B}$ can be seen as generalized Sylvester equations and solved by efficient solvers, for instance, Hessenberg-Schur or Bartels-Stewart algorithms; see [37] for a full overview. The updates for $\boldsymbol{C}$ and $\widetilde{\boldsymbol{C}}_M$ are solved using normal equations. Please refer to Appendix A for a full derivation.

The computational cost per iteration of BTD-Var can be decomposed as follows:

- $\mathcal{O}(I^3 + J^3 + L^3R^3)$ for solving the Sylvester equations;
- $\mathcal{O}(IJK_MR + I_HJ_HKR)$ for computing the right-hand sides in the least squares subproblems.

[3]As in previous works [14]–[18], we further consider that $\lambda = 1$ in our experiments.

Algorithm 1: BTD-Var

input : $\mathcal{Y}_H$, $\mathcal{Y}_M$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\widetilde{\boldsymbol{C}}_M$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$; R, L, `iter`
output: $\mathcal{Z} \in \mathbb{R}^{I\times J\times K}$, $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3 \in \mathbb{R}^{I\times J\times K_M}$
for $m \in \{1,\ldots,$ `iter`$\}$ **do**

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \ \|\boldsymbol{Y}_H^{(1)} - (\boldsymbol{C} \odot_p \boldsymbol{P}_2\boldsymbol{B})\,\boldsymbol{A}^\mathsf{T}\boldsymbol{P}_1^\mathsf{T}\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(1)} - \left(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{B}\right)\boldsymbol{A}^\mathsf{T}\|_F^2,$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \ \|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot_p \boldsymbol{P}_1\boldsymbol{A})\,\boldsymbol{B}^\mathsf{T}\boldsymbol{P}_2^\mathsf{T}\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(2)} - \left(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{A}\right)\boldsymbol{B}^\mathsf{T}\|_F^2,$$

$$\boldsymbol{S}_r \leftarrow \mathrm{vec}\{\boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T}\} \text{ for } r \in \{1,\ldots,R\},$$

$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \ \|\boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S}\boldsymbol{C}^\mathsf{T}\|_F^2,$$

$$\widetilde{\boldsymbol{C}}_M \leftarrow \arg\min_{\widetilde{\boldsymbol{C}}_M} \ \lambda\|\boldsymbol{Y}_H^{(3)} - \boldsymbol{S}\widetilde{\boldsymbol{C}}_M^\mathsf{T}\|_F^2.$$

  for $r \in \{1,\ldots,R\}$ **do**

$$\boldsymbol{c}_r = \boldsymbol{c}_r/\|\boldsymbol{c}_r\|, \quad \widetilde{\boldsymbol{c}}_{M,r} = \widetilde{\boldsymbol{c}}_{M,r}/\|\widetilde{\boldsymbol{c}}_{M,r}\|.$$

  end

$$\boldsymbol{Z}^{(3)} = \boldsymbol{S}\boldsymbol{C}^\mathsf{T},\ (\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3)^{(3)} = \boldsymbol{S}(\widetilde{\boldsymbol{C}}_M - \boldsymbol{P}_3\boldsymbol{C})^\mathsf{T}.$$

end

Algorithm 2: CNN-BTD-Var

input : $\mathcal{Y}_H$, $\mathcal{Y}_M$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\widetilde{\boldsymbol{C}}_M$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$; R, L, `iter`
output: $\boldsymbol{S} \in \mathbb{R}^{IJ\times R}$, $\boldsymbol{C} \in \mathbb{R}^{K\times R}$, $\widetilde{\boldsymbol{C}}_M \in \mathbb{R}^{K_M\times R}$, $\mathcal{Z} \in \mathbb{R}^{I\times J\times K}$, $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3 \in \mathbb{R}^{I\times J\times K_M}$
for $m \in \{1,\ldots,$ `iter`$\}$ **do**

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \ \|\boldsymbol{Y}_H^{(1)} - (\boldsymbol{C} \odot_p \boldsymbol{P}_2\boldsymbol{B})\,\boldsymbol{A}^\mathsf{T}\boldsymbol{P}_1^\mathsf{T}\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(1)} - \left(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{B}\right)\boldsymbol{A}^\mathsf{T}\|_F^2,$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \ \|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot_p \boldsymbol{P}_1\boldsymbol{A})\,\boldsymbol{B}^\mathsf{T}\boldsymbol{P}_2^\mathsf{T}\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(2)} - \left(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{A}\right)\boldsymbol{B}^\mathsf{T}\|_F^2,$$

$$\boldsymbol{S}_r \leftarrow \|\boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T} - \boldsymbol{S}_r\|_F^2 + \iota_+(\boldsymbol{S}_r),$$

$$\boldsymbol{C} \leftarrow \|\boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S}\boldsymbol{C}^\mathsf{T}\|_F^2 + \iota_+(\boldsymbol{C}),$$

$$\widetilde{\boldsymbol{C}}_M \leftarrow \lambda\|\boldsymbol{Y}_M^{(3)} - \boldsymbol{S}\widetilde{\boldsymbol{C}}_M^\mathsf{T}\|_F^2 + \iota_+(\widetilde{\boldsymbol{C}}_M).$$

  for $r \in \{1,\ldots,R\}$ **do**

$$\boldsymbol{c}_r = \boldsymbol{c}_r/\|\boldsymbol{c}_r\|, \quad \widetilde{\boldsymbol{c}}_{M,r} = \widetilde{\boldsymbol{c}}_{M,r}/\|\widetilde{\boldsymbol{c}}_{M,r}\|.$$

  end

$$\boldsymbol{Z}^{(3)} = \boldsymbol{S}\boldsymbol{C}^\mathsf{T},\ (\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3)^{(3)} = \boldsymbol{S}(\widetilde{\boldsymbol{C}}_M - \boldsymbol{P}_3\boldsymbol{C})^\mathsf{T}.$$

end

B. Constrained optimization

Although BTD-Var allows for reconstruction of $\mathcal{Z}$ and $\boldsymbol{\Psi}$, it is not guaranteed that its result can be used for the HU task. To that end, non-negativity constraints must be imposed on factors $\boldsymbol{C}$ and $\widetilde{\boldsymbol{C}}_M$ to provide them with physical meaning. Moreover, contrary to [17], we also impose non-negativity on $\{\boldsymbol{S}_r\}_{r=1}^R$, rather than on the individual factors $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$. This way, $\boldsymbol{c}_r$ and $\widetilde{\boldsymbol{c}}_{M,r}$ (resp. $\boldsymbol{S}_r$) can be seen as spectral signatures (resp. abundance maps) of the underlying SRI $\mathcal{Z}$ and MSI $\mathcal{Y}_M$.

The resulting constrained optimization problem is:

$$\underset{\boldsymbol{A},\boldsymbol{B},\{\boldsymbol{S}_r\}_{r=1}^R,\boldsymbol{C},\widetilde{\boldsymbol{C}}_M}{\text{minimize}} \ \mathcal{J}(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\widetilde{\boldsymbol{C}}_M) \tag{17}$$

$$\text{s. to } \{\boldsymbol{S}_r = \boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T}\}_{r=1}^R \geq \boldsymbol{0}, \boldsymbol{C} \geq \boldsymbol{0}, \widetilde{\boldsymbol{C}}_M \geq \boldsymbol{0}, \tag{18}$$

While the updates for $\boldsymbol{A}$ and $\boldsymbol{B}$ are the same as in BTD-Var, the other updates are constrained quadratic programs. The non-negativity constraints can be handled by using alternating direction method of multipliers (ADMM) [38], [39]. As in [39], a non-negativity constraint is relaxed by considering the surrogate $\iota_+(\cdot)$ (see Appendix B). Algorithm 2 presents the optimization framework for (17)–(18).

The computational cost per-iteration of CNN-BTD-Var is:

- $\mathcal{O}(I^3 + J^3 + L^3R^3)$ for solving $\boldsymbol{A}$ and $\boldsymbol{B}$;
- $\mathcal{O}(IJK_MR + I_HJ_HKR)$ for computing the right-hand sides in the least squares subproblems.

C. Initialization

Many options are available to initialize the LL1 factors. Here, as suggested in [29, Theorem 4.1], we initialize the $\boldsymbol{A}$ and $\boldsymbol{B}$ factors by generalized eigenvalue decomposition (GEVD) of the matrix pencil $\left((\mathcal{Y}_M)_{:,:,1}^\mathsf{T}, (\mathcal{Y}_M)_{:,:,2}^\mathsf{T}\right)$ (see [40], [41]), using the `lll_gevd` function of TensorLab [42]. The $\boldsymbol{C}$ and $\widetilde{\boldsymbol{C}}_M$ factors are recovered by solving least-squares problems. We combine these steps in an algebraic algorithm called BTDRec (Algorithm 3), echoing the initialization algorithm in [14] (called TenRec):

Algorithm 3: BTDRec

input : $\mathcal{Y}_H$, $\mathcal{Y}_M$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$; R, L
output: $\boldsymbol{A} \in \mathbb{R}^{I\times RL}$, $\boldsymbol{B} \in \mathbb{R}^{J\times RL}$, $\boldsymbol{C} \in \mathbb{R}^{K\times R}$, $\widetilde{\boldsymbol{C}}_M \in \mathbb{R}^{K_M\times R}$

1. $\boldsymbol{A}, \boldsymbol{B} \overset{\text{LL1}}{\approx} \mathcal{Y}_M$;
2. $\boldsymbol{S}_r = \boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T}$ for $r \in \{1,\ldots,R\}$;
3. $\boldsymbol{C}^\mathsf{T} = ((\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S})^\dagger\, \boldsymbol{Y}_H^{(3)}$;
4. $\widetilde{\boldsymbol{C}}_M^\mathsf{T} = \boldsymbol{S}^\dagger \boldsymbol{Y}_M^{(3)}$.

V. Experiments for image recovery

All simulations were run on a MacBook Pro with 2.3 GHz Intel Core i5 and 16GB RAM. The code was implemented in MATLAB. For basic tensor operations we used TensorLab 3.0 [42]. The code is implemented in MATLAB and available online at https://github.com/cprevost4/LL1_HSR_HU.

A. Metrics

We compared the groundtruth SRI $\mathcal{Z}$ with the recovered SRI $\widehat{\mathcal{Z}}$ obtained by the algorithms. The main performance metric used in comparisons was the *reconstruction Signal-to-Noise ratio* (R-SNR):

$$\text{R-SNR} = 10\log_{10}\left(\frac{\|\mathcal{Z}\|_F^2}{\|\widehat{\mathcal{Z}} - \mathcal{Z}\|_F^2}\right). \tag{19}$$

In addition to R-SNR, we considered different metrics described below:

$$\text{CC} = \frac{1}{IJK}\left(\sum_{k=1}^K \rho\left(\mathcal{Z}_{:,:,k}, \widehat{\mathcal{Z}}_{:,:,k}\right)\right), \tag{20}$$

where $\rho(\cdot,\cdot)$ is the Pearson correlation coefficient between the estimated and original spectral slices;

$$\text{ERGAS} = \frac{100}{d}\sqrt{\frac{1}{IJK}\sum_{k=1}^K \frac{\|\widehat{\mathcal{Z}}_{:,:,k} - \mathcal{Z}_{:,:,k}\|_F^2}{\mu_k^2}}, \tag{21}$$

where μ_k^2 is the mean value of $\widehat{\boldsymbol{\mathcal{Z}}}_{:,:,k}$. ERGAS represents the relative dimensionless global error between the SRI and the estimate, which is the root mean-square error averaged by the size of the SRI. We also used Spectral Angle Distance (SAD):

$$\text{SAD} = \frac{1}{R}\sum_{r=1}^{R} \arccos\left(\frac{\boldsymbol{c}_r^{\mathsf{T}}\widehat{\boldsymbol{c}}_r}{\|\boldsymbol{c}_r\|_2\|\widehat{\boldsymbol{c}}_r\|_2}\right), \tag{22}$$

which computes the spectral angle distance between original and estimated spectra, and can be used to assess unmixing performance as well. Finally, we considered the computational time for each algorithm, given by the `tic` and `toc` functions of MATLAB.

B. Degradation model

The HSI was obtained by spatial degradation of $\boldsymbol{\mathcal{Z}}$ by $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, *i.e.*, the SRI $\boldsymbol{\mathcal{Z}}$ and the MSI $\boldsymbol{\mathcal{Y}}_M$ represent images of the same scene acquired on board of different missions, and $\widetilde{\boldsymbol{\mathcal{Z}}}$ is unknown. The spectral bands of $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{Y}}_M$ were normalized such that the 0.999 intensity quantile corresponded to a value of 1. Afterwards, the SRI $\boldsymbol{\mathcal{Z}}$ was denoised (as described in [43]) to yield the high-SNR reference image [5].

We also conducted experiments in a "no-variability" scenario, *i.e.* we consider that the HSI and MSI were obtained by spatial (resp. spectral) degradation of the same SRI $\boldsymbol{\mathcal{Z}}$.

For spatial degradation, we followed the commonly used Wald's protocol [30]. The matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ were computed with a separable Gaussian blurring kernel of size $q = 9$. Downsampling was performed along each spatial dimension with a ratio $d = 4$ between (I, J) and (I_H, J_H), as in previous works [14]–[18]. Refer to Appendix C for more details on the construction of $\boldsymbol{P}_1$, $\boldsymbol{P}_2$. White Gaussian noise with 30dB SNR was added to the HSI and MSI.

For the spectral degradation matrix $\boldsymbol{P}_3$, we used the SRF of two multispectral instruments[4]. For images with spectral variability, the Sentinel-2 sensors span the electromagnetic spectrum from 412nm to 2022nm and produce a 10-band MSI corresponding to the wavelengths 433–453nm (atmospheric correction), 458–522nm (soil, vegetation), 543–577nm (green peak), 650–680nm (maximum chlorophyll absorption), 698–712nm (red edge), 733–747nm (red edge), 773–793nm (leaf area index, edge of NIR), 785–900nm (leaf area index), 855–875nm (NIR plateau), 935–955nm (water vapour absorption). The LANDSAT sensor spans the spectrum from 400nm to 2500nm for the HSI and produces a 6-band MSI corresponding to wavelengths 450–520nm (blue), 520–600nm (green), 630–690nm (red), 760–900nm (near-IR), 1550–1750nm (shortwave-IR) and 2050–2350nm (shortwave-IR2). This spectral response is used for semi-real images without spectral variability. The spectral degradation matrix $\boldsymbol{P}_3$ is a selection-weighting matrix that selects the common spectral bands of the SRI $\widetilde{\boldsymbol{\mathcal{Z}}}$ and the MSI.

C. Recovery of the SRI and variability tensor

In this subsection, we assess the performances of Algorithm 1 (BTD-Var) and Algorithm 2 (CNN-BTD-Var) for reconstruction of the SRI $\boldsymbol{\mathcal{Z}}$ and degraded variability tensor $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$. We ran our algorithms with 20 outer iterations at most and 5 ADMM iterations for CNN-BTD-Var. For initialization, out of 20 trials of BTDRec we picked the one that provided the best reconstruction of the HSI and MSI.

For the SRI $\boldsymbol{\mathcal{Z}}$, we compared our results to matrix-based approaches, including HySure [8], CNMF [6] and GLP-HS [46]. We also considered tensor factorization methods, namely STEREO [14] for CP decomposition, SCOTT [16] for Tucker and CNN-BTD [17], which is a coupled LL1-based algorithm that does not account for spectral variability. Finally, we compared our approach to matrix and tensor methods accounting for variability, namely FuVar [24] (a matrix-based algorithm based on the GLMM), CT-STAR and CB-STAR [26], which are tensor approaches based on multilinear decomposition accounting for both spatial and spectral variability. It must be noticed that except for CNMF, the baseline algorithms are unable to perform the unmixing task. For Hysure, CNMF, GLP-HS and FuVar, we chose the ranks and regularization parameters according to the original works [8], [6], [24].

For reconstruction of $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$, we compared the results of our algorithms with those of CT-STAR and CB-STAR. We assessed the performance by computing R-SNR, CC, ERGAS and SAD for each algorithm.

1) Lake Tahoe: The first dataset we considered was Lake Tahoe with $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{100\times 80\times 173}$. The SRI $\boldsymbol{\mathcal{Z}}$ and MSI $\boldsymbol{\mathcal{Y}}_M$ were respectively acquired on 2014-10-04 and 2017-10-24 by the Sentinel-2A sensor, resulting in high variability in the crops and lake areas). We ran STEREO with $F = 30$ and 10 iterations, and SCOTT with $R = (40, 40, 7)$ as in [26]. We ran CT-STAR with ranks $(18, 15, 10)$,$(3, 3, 1)$, and CB-STAR with ranks $(20, 20, 9)$,$(20, 20, 4)$. For our algorithms, as well as for CNN-BTD, we chose $R = 3$, $L = 20$ and $\lambda = 1$.

Tables I and II display the reconstruction metrics and computation time for $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$ and all considered algorithms. The two best results of each column are shown in bold.

TABLE I
RECONSTRUCTION METRICS FOR $\boldsymbol{\mathcal{Z}}$, LAKE TAHOE DATASET

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	15.0674	**0.93829**	9.6873	5.2145	1.4818
CNN-BTD-Var	**15.1093**	0.93574	8.0109	**5.1817**	0.9025
STEREO	5.8368	0.75957	30.7346	15.2801	1.2148
SCOTT	1.918	0.50379	47.1781	23.3815	**0.14701**
CNN-BTD	6.0332	0.80003	27.7993	14.9491	1.2826
CNMF	12.1314	0.87494	9.2422	7.2804	1.7442
GLP-HS	11.7862	0.87408	11.6106	7.6011	4.507
HySure	9.2687	0.81256	12.8228	10.1511	7.2761
FuVar	14.54	0.92498	**6.7013**	5.528	761.3932
CT-STAR	11.7676	0.87843	13.3433	7.6236	**0.20849**
CB-STAR	**19.2413**	**0.97539**	**6.4649**	**3.2231**	8.3597

TABLE II
RECONSTRUCTION METRICS FOR $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$, LAKE TAHOE DATASET

Algorithm	R-SNR	CC	SAD	ERGAS
BTD-Var	13.7482	0.85583	**14.8728**	12.121
CNN-BTD-Var	**13.7643**	**0.88335**	18.9519	**10.7103**
CT-STAR	11.4131	0.84542	17.7857	12.8223
CB-STAR	**16.6599**	**0.94161**	**10.4442**	**7.8569**

We can see that algorithms accounting for variability provided the best reconstruction metrics: in particular, the high performance of CB-STAR resulted from the fact that the algorithm takes into account both spatial and spectral variabilities. BTD-Var and CNN-BTD-Var provided metrics comparable to those of FuVar, but with lower computation time.

[4]available for download at [44] and [45].

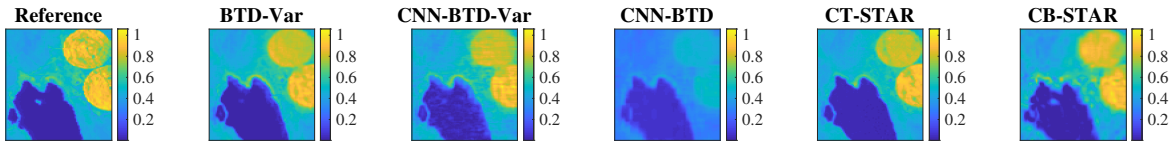

Fig. 1. Single spectral band of the SRI, Lake Tahoe dataset

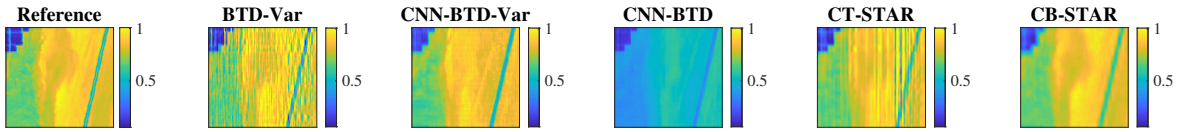

Fig. 2. Single spectral band of the SRI, Ivanpah Playa dataset

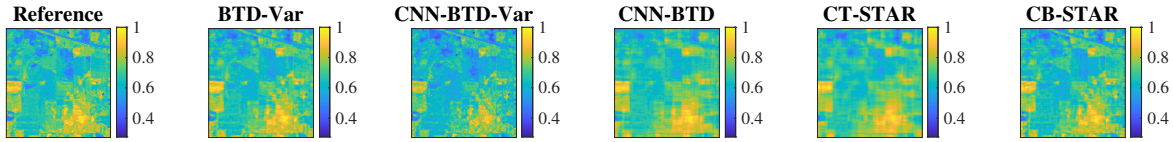

Fig. 3. Single spectral band of the SRI, Indian Pines dataset

Among the matrix-based approaches, CNMF showed the best reconstruction performance. Finally, state-of-the art tensor-based approaches, although fast, yielded worse reconstruction metrics than the aforementioned methods. This was due to the fact that they do not consider any kind of variability.

CB-STAR also provided the best metrics for reconstruction of $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$. However, its computation time was large. The proposed algorithms showed competitive metrics and even slightly outperformed baseline methods in terms of CC, but with slightly higher computation time.

In addition, we plot in Figure 1 the 40th spectral band of the reference SRI, as well as the estimated SRI for our algorithms. The proposed approaches recovered the SRI spectral band accurately.

2) Ivanpah Playa: We also considered the Ivanpah Playa dataset with $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{80\times128\times173}$ and large acquisition time difference: the SRI and MSI were acquired on 2015-10-26 and 2017-12-17 respectively, by the Sentinel-2 sensor. We ran STEREO with $F = 10$ and 10 iterations and SCOTT with $R = (30, 30, 10)$. We ran CT-STAR with ranks $(10, 15, 8)$,$(3, 3, 2)$, and CB-STAR with ranks $(40, 40, 4)$,$(40, 40, 5)$. For our algorithms, as well as for CNN-BTD, we chose $R = 4$ and $L = 18$.

Tables III and IV show the reconstruction metrics and computation time for various algorithms.

The best metrics were provided by CB-STAR, then CNN-BTD-Var. BTD-Var had a performance comparable to that of GLP-HS for reconstruction of $\boldsymbol{\mathcal{Z}}$. Its performance was comparable to that of CT-STAR for $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$. For this dataset as well, the proposed algorithms were faster than some other algorithms, including CNMF, CB-STAR, and FuVar. Other matrix-based approaches also provided satisfying reconstruction. However, STEREO, SCOTT and CNN-BTD provided the worst reconstruction metrics.

In Figure 2 we plot the 40th spectral band of the reference SRI, as well as the estimated SRI for our algorithms, CNN-BTD, CNMF and CB-STAR for comparison. For this dataset, we can see that CNN-BTD-Var yielded a better SRI spectral band reconstruction than BTD-Var.

TABLE III
RECONSTRUCTION METRICS FOR $\boldsymbol{\mathcal{Z}}$, IVANPAH PLAYA DATASET

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	19.4098	0.86314	2.4404	2.6771	2.1568
CNN-BTD-Var	**22.7305**	**0.92074**	2.6247	**1.8919**	1.386
STEREO	6.0987	0.76283	29.0278	12.6747	0.93975
SCOTT	2.4445	0.34257	47.9598	19.372	**0.2645**
CNN-BTD	5.7515	0.33492	28.7006	13.1899	11.8775
CNMF	21.6059	0.90114	**1.3019**	2.1138	2.6656
GLP-HS	19.433	0.86261	3.3413	2.697	5.9218
HySure	18.4551	0.85218	3.3249	3.0653	10.4606
FuVar	22.0332	0.90354	1.5062	2.0189	526.1659
CT-STAR	21.1186	0.88849	1.9424	2.2386	**0.15373**
CB-STAR	**25.7174**	**0.96003**	**1.3269**	**1.3228**	8.2923

TABLE IV
RECONSTRUCTION METRICS FOR $\boldsymbol{\Psi} \bullet_3 \boldsymbol{P}_3$, LAKE TAHOE DATASET

Algorithm	R-SNR	CC	SAD	ERGAS
BTD-Var	19.0156	0.68664	3.3129	43.8732
CNN-BTD-Var	**21.75**	**0.80652**	**2.0192**	**25.1069**
CT-STAR	19.3597	0.73396	2.1977	33.853
CB-STAR	**23.4888**	**0.90832**	**1.1567**	**16.9815**

D. Recovery without variability

In this subsection, we assess recovery performance for the SRI $\boldsymbol{\mathcal{Z}}$. We consider a "no-variability" scenario, *i.e.* the HSI and MSI are both degraded versions of $\boldsymbol{\mathcal{Z}}$. The dataset we consider is Indian Pines, where $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{144\times144\times200}$ is degraded by a LANDSAT sensor for the MSI and a downsampling ratio $d = 4$ for the HSI. We ran STEREO with $F = 50$, SCOTT with $R = (40, 40, 6)$ and LL1-based algorithms with $R = 6$ and $L = 13$. We tuned the other algorithms according to original works. The reconstruction metrics for the SRI are presented in Table V.

TABLE V
RECONSTRUCTION METRICS FOR $\mathcal{Z}$, INDIAN PINES DATASET

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	**28.0511**	**0.87064**	**1.9204**	**0.98804**	8.3579
CNN-BTD-Var	26.2721	0.8327	2.2515	1.2191	3.176
STEREO	**27.69**	0.86669	1.9461	**0.99959**	1.8564
SCOTT	26.2451	0.86196	2.2694	1.1208	**0.21087**
CNN-BTD	25.2263	0.80949	2.5035	1.3497	24.5326
CNMF	27.2552	0.83978	**1.9502**	1.2056	8.2147
GLP-HS	26.2837	0.83813	2.2794	1.2918	14.2957
HySure	20.4281	0.66661	4.4916	2.5723	25.2202
CT-STAR	24.0398	0.84385	2.4839	1.3151	**0.16528**
CB-STAR	26.5216	**0.86749**	2.1265	1.0556	3.6761

The best reconstruction metrics were generally provided by STEREO and BTD-Var. CNN-BTD-Var had performance comparable to that of SCOTT, and computation time comparable to that of CB-STAR. The slightly lower performance of constrained algorithms accounting for variability can be explained by the use of more flexible models. In this specific scenario, other methods based on a more restrictive model fit the data more tightly. Nonetheless, algorithms accounting for variability offer competitive performance in the "no-variability" case. However, their computation time is usually higher than that of state-of-the-art tensor approaches. In Figure 3 we plot the 40th spectral band of the reference and estimated SRI.

VI. BLIND UNMIXING EXPERIMENTS

A. Experiments setup

In this section, we assessed the performance of CNN-BTD-Var for hyperspectral unmixing on synthetic datasets, and semi-real examples from Section V.

We compared our results with those of CNMF [6] initialized by VCA [47]. We also considered traditional unmixing algorithms: accelerated multiplicative algorithm (MU-Acc) [48] and BMDR-ADMM [3] (enforcing minimum dispersion constraint on the spectra, and sum-to-one on the abundance maps). We ran these algorithms on the recovered SRI $\widehat{\mathcal{Z}}$ obtained from CB-STAR, which gave the best reconstruction metrics in the previous section. For these algorithms, we chose the parameters according to the original works. We used the actual number of materials as R. Since other fusion algorithms such as FuVar, CT-STAR, CB-STAR are not designed for the HU task, they were not directly included in this comparison.

For each dataset, we compared reference abundance maps and spectral signatures to those obtained by the algorithms.

B. Hyperspectral unmixing with exact LL1 model

We first assessed the unmixing performance of our algorithm in a controlled environment with synthetic datasets. That is, we tested our approach in the case where the SRI and variability tensor admit an exact LL1-BTD. Although these datasets do resemble real spectral images, they allow us to assess unmixing performance of our algorithm in a case where the uniqueness conditions for the NMF (see [11], [12]) are not fulfilled.

1) Generating synthetic datasets: We considered $R = 3$ spectral signatures $\boldsymbol{c}_r$ ($r \in \{1, \ldots, R\}$) obtained from the Jasper Ridge reference data[5], corresponding to vegetation, soil and road materials. The SRI $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I \times J \times K}$ ($I = J = 90$, $K = 173$) was split into LR equal blocks in the spatial dimensions, with $L = 3$.

Each abundance map $\boldsymbol{S}_r$ ($r \in \{1, \ldots, R\}$) was a block matrix with $\frac{I}{L} \times \frac{J}{L}$ blocks.

We generated a multiplicative variability matrix $\boldsymbol{\psi}_{\text{multi}}$ with random real entries drawn from the standard uniform distribution in the open interval $[0.9, 1.1]$. We then computed the altered spectra $\widetilde{\boldsymbol{C}} = \boldsymbol{\psi}_{\text{multi}} \boxdot \boldsymbol{C}$. The variability matrix $\boldsymbol{\psi}$ that we aimed at recovering was obtained as $\boldsymbol{\psi} = \widetilde{\boldsymbol{C}} - \boldsymbol{C}$ so that it had zero mean.

Formally, we computed the high-resolution tensors as

$$\boldsymbol{\mathcal{Z}} = \sum_{r=1}^{R} \boldsymbol{S}_r \otimes \boldsymbol{c}_r, \quad \boldsymbol{\Psi} = \sum_{r=1}^{R} \boldsymbol{S}_r \otimes \boldsymbol{\psi}_r, \quad \widetilde{\boldsymbol{\mathcal{Z}}} = \boldsymbol{\mathcal{Z}} + \boldsymbol{\Psi}.$$

The HSI and MSI were obtained by degradation of the SRIs according to model (4). For $\boldsymbol{P}_1 = \boldsymbol{P}_2$, we had $q = 9$ and $d = 3$ so that $I_H = J_H = 30$. For $\boldsymbol{P}_3$, we chose the SRF matrix of the Sentinel-2 MS sensor, which led to $K_M = 10$.

2) Separable example: In the first example, we generated a dataset for which the pure pixel assumption is valid. Thus in each $\frac{I}{L} \times \frac{J}{L}$ block, at most one material was active, as indicated by the numerals in the parcel map shown in Table VI. Each block in the parcel map was a patch composed of

TABLE VI
PARCEL MAP FOR THE FIRST SYNTHETIC DATASET

1	2	**3**
3	1	2
2	**3**	1

entries equal to one. The abundance maps in this case mimic agricultural fields. This is a case for which the NMF under minimal volume constraint is unique [49], [50], [51]. Only unconstrained NMF is not unique.

We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The spectral signatures $\boldsymbol{c}_r$ and abundance maps $\boldsymbol{S}_r$ are shown in Figures 4 and 5, respectively.

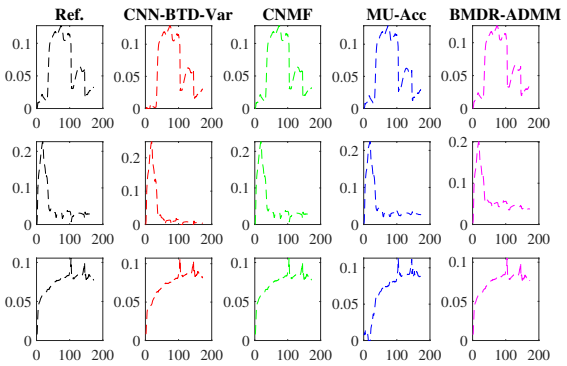

Fig. 4. Reference and estimated spectra, synthetic dataset 1

We can see that all spectra and abundance maps were recovered accurately by CNN-BTD-Var, with visual quality comparable to that of CNMF and BMDR-ADMM. We also see some artifacts in the abundance maps recovered by CNMF. Moreover, MU-Acc did not estimate all abundance maps correctly for this example.

[5]Available for download here.

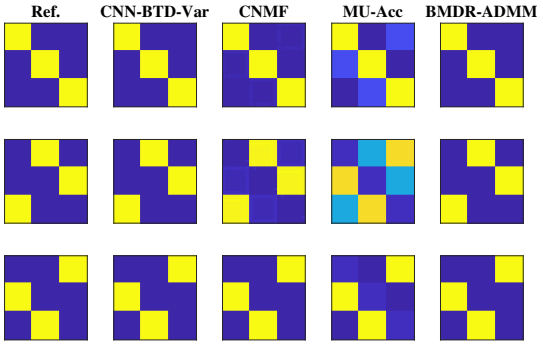

Fig. 5. Reference and estimated abundance maps, synthetic dataset 1

3) Synthetic example with non-identifiable NMF: In this second example, we aim at designing an example where the separability (or pure pixel) condition [11], [12] is not fulfilled. This results in the traditional NMF being non-identifiable. This is in fact a highly mixed situation for which all existing approaches will fail. Thus, we expect that traditional unmixing algorithms fail at performing the HU task on this dataset. However, the conditions in Theorem III.1 are satisfied, which makes the LL1 factors unique up to permutation and scaling ambiguities. The abundance maps $\boldsymbol{S}_r$ are designed as follows:

$$\boldsymbol{S}_1 = \frac{1}{12}\begin{bmatrix} 5 & 7 & 6 \\ 7 & 3 & 5 \\ 3 & 0 & 0 \end{bmatrix} \boxtimes \boldsymbol{H}, \quad \boldsymbol{S}_2 = \frac{1}{12}\begin{bmatrix} 7 & 5 & 3 \\ 0 & 6 & 0 \\ 3 & 5 & 7 \end{bmatrix} \boxtimes \boldsymbol{H},$$

$$\boldsymbol{S}_3 = \frac{1}{12}\begin{bmatrix} 0 & 0 & 3 \\ 5 & 3 & 7 \\ 6 & 7 & 5 \end{bmatrix} \boxtimes \boldsymbol{H},$$

with $\boldsymbol{H}$ a Gaussian of size 30×30 with standard deviation $\sigma = 5$. From the above abundance maps, we can see that the pure pixel assumption is not valid.

We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The spectral signatures $\boldsymbol{c}_r$ and abundance maps $\boldsymbol{S}_r$ are shown in Figures 6 and 7, respectively.

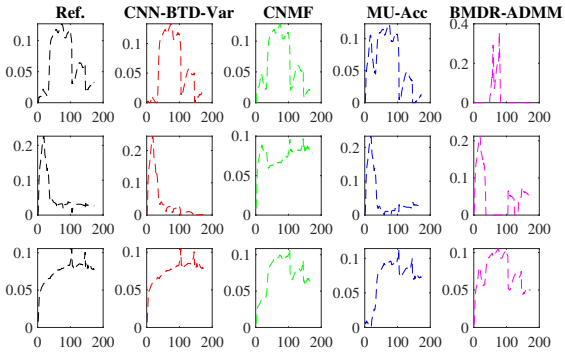

Fig. 6. Reference and estimated spectra, synthetic dataset 2

We can see that the spectral signatures are best reconstructed by CNN-BTD-Var, although CNMF only reconstructs the first spectrum correcly. Moreover, only CNN-BTD-Var provides reasonable estimates of the abundance maps, while other algorithms fail.

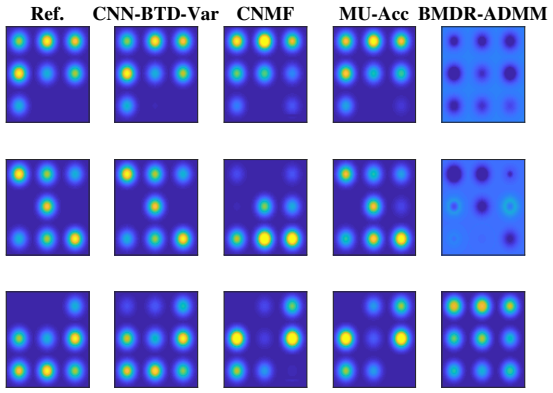

Fig. 7. Reference and estimated abundance maps, synthetic dataset 2

C. Hyperspectral unmixing for semi-real datasets

In this subsection, we assess HU performance of CNN-BTD-Var for semi-real datasets Lake Tahoe and Ivanpah Playa. For the two considered datasets, we followed the same degradation model as in Section V. For these experiments, the endmembers and abundance maps underlying $\boldsymbol{\mathcal{Z}}$ are unknown: as a result, we chose as reference the spectra and abundance maps selected manually from the SRI $\boldsymbol{\mathcal{Z}}$. The obtained abundance maps had very close correspondence with visual features in the image. The columns of the spectra and abundance maps were rescaled with unit norm for comparison.

1) Lake Tahoe: We first performed HU on the Lake Tahoe dataset with $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{100\times 80\times 173}$. This dataset is mainly composed of $R = 3$ materials: water (lake), soil and vegetation (under the form of crop circles). As a result, we chose $R = 3$ and $L = 18$ as in the previous subsection. We compared our algorithm with CNMF, MU-Acc and BMDR-ADMM with $R = 3$.

On Figures 8 and 9, we plot the estimated spectra and abundance maps.

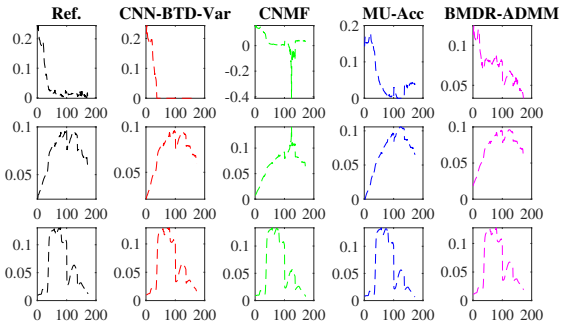

Fig. 8. Reference and estimated spectra, Lake Tahoe dataset

We notice that the proposed approach estimated the spectra accurately. However, CNMF failed at recovering the water spectrum. The abundance maps displayed in Figure 9 allow for identification of the areas corresponding to different materials, although with lower resolution than other methods. Additionally, the abundance maps recovered by CNN-BTD-Var seem to be low-rank. We can see that CNMF and MU-Acc did not recover the water abundance map correctly.

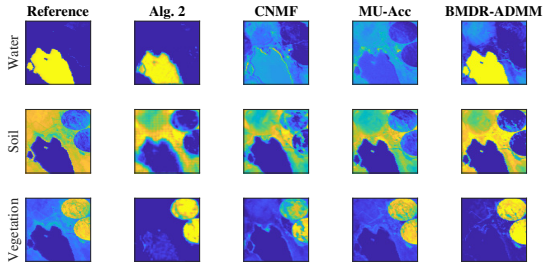

Fig. 9. Reference and estimated abundance maps, Lake Tahoe dataset

Additionally, on Figure 10, we plot the reference and estimated $\widetilde{\boldsymbol{C}}_M$ and $\boldsymbol{P}_3\boldsymbol{\psi} = \widetilde{\boldsymbol{C}}_M - \boldsymbol{P}_3\boldsymbol{C}$ obtained from CNN-BTD-Var.

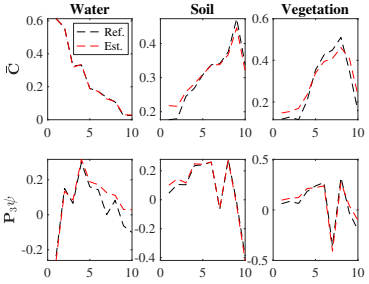

Fig. 10. Reference and estimated $\widetilde{\boldsymbol{C}}_M$ and $\boldsymbol{P}_3\boldsymbol{\psi}$, Lake Tahoe dataset

In Figure 10, we can see that the water spectrum presents high variability for the first MSI spectral band, which corresponds to the blue region. For the vegetation and soil spectra, the most variability can be found at spectral bands corresponding to the green and orange-red wavelengths. Moreover, CNN-BTD-Var recovers the reference $\widetilde{\boldsymbol{C}}_M$ and $\boldsymbol{P}_3\boldsymbol{\psi}$ with a small discrepancy.

2) Ivanpah Playa: Next, we considered the Ivanpah Playa dataset. This dataset is composed of $R = 4$ materials: solar panels, dark sand, yellow sand and road. We ran CNN-BTD-Var with $R = 4$, $L = 18$, and compared the results to other baseline algorithms with $R = 4$.

In Figures 11 and 12, we plot the reference and estimated spectra and abundance maps.

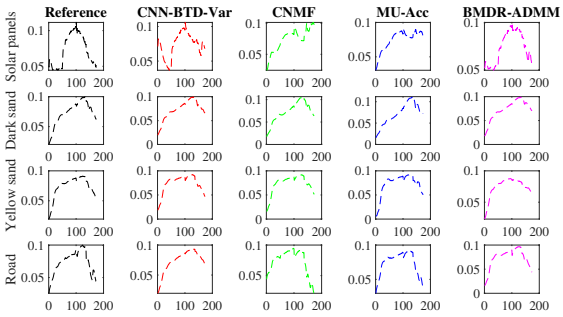

Fig. 11. Reference and estimated spectra, Ivanpah Playa dataset

One difficulty that arises for the HU task is the important similarity between the reference spectra (in particular, dark and yellow sand, road materials). This leads to almost colinear columns in $\boldsymbol{C}$, which may lead to poorer performance. This issue is particularly visible in Figure 13 with estimated $\widetilde{\boldsymbol{C}}_M$ and $\boldsymbol{P}_3\boldsymbol{\psi}$.

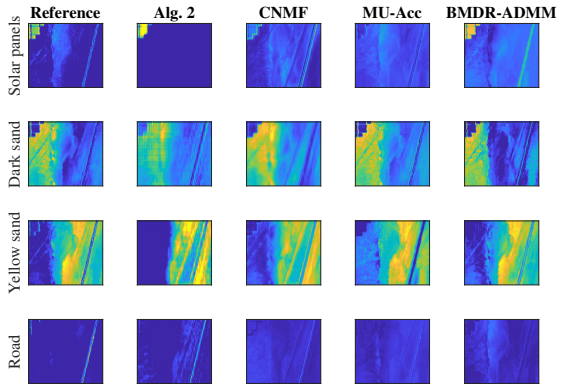

Fig. 12. Reference and estimated abundance maps, Ivanpah Playa dataset

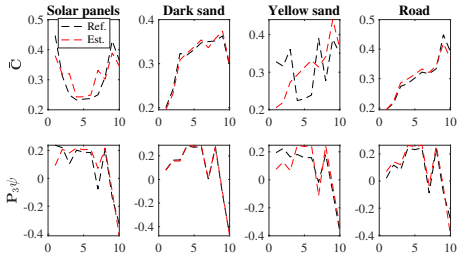

Fig. 13. Reference and estimated $\widetilde{\boldsymbol{C}}_M$ and $\boldsymbol{P}_3\boldsymbol{\psi}$, Ivanpah Playa dataset

In Figure 13, we can see that high variability can be found for the red and near-infrared spectral bands for all materials.

Despite this difficulty, CNN-BTD-Var recovered the solar panels and road abundance maps best, while the yellow sand map was slightly better for BMDR-ADMM. For this example, the low-rank assumption for abundance maps is reasonable: see that corresponding to solar panels. This assumption allows for better visual reconstruction of this abundance map. Contrary to other algorithms, all spectra are correctly recovered.

VII. Conclusion

In this paper, we proposed new algorithms for solving the HSR problem with variable images, using an LL1-BTD model. First, we showed that in the presence of variability, previous tensor models fail at recovering the SRI, since they do not account for spectral or spatial variability. Our approach allows to recover the SRI accurately for the considered datasets, as well as the degraded variability tensor.

An appropriate choice of ranks also allows our algorithms to estimate underlying spectra and abundance maps of the SRI, with performance comparable to those of algorithms that perform unmixing on the SRI directly. Non-negativity priors allow the low-rank factors of our model to be interpretable, without having a high negative impact on the computation time.

Appendix A
Unconstrained Factor Updates

The unconstrained least squares program for $\boldsymbol{A}$ can be seen as a generalized Sylvester equation of the form $\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_2 +$

$\boldsymbol{X}_3\boldsymbol{A}\boldsymbol{X}_4 = \boldsymbol{X}_5$, with

$$\begin{aligned}
&\boldsymbol{X}_1 = \boldsymbol{P}_1^\mathsf{T}\boldsymbol{P}_1, \quad \boldsymbol{X}_2 = (\boldsymbol{C} \odot_p \boldsymbol{P}_2\boldsymbol{B})^\mathsf{T}(\boldsymbol{C} \odot_p \boldsymbol{P}_2\boldsymbol{B}),\\
&\boldsymbol{X}_3 = \lambda\boldsymbol{I}_I, \qquad \boldsymbol{X}_4 = (\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{B})^\mathsf{T}(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{B}),\\
&\boldsymbol{X}_5 = \boldsymbol{P}_1^\mathsf{T}(\boldsymbol{Y}_H^{(1)})^\mathsf{T}(\boldsymbol{C} \odot_p \boldsymbol{P}_2\boldsymbol{B}) + \lambda(\boldsymbol{Y}_M^{(1)})^\mathsf{T}(\widetilde{\boldsymbol{C}}_M \odot_p \boldsymbol{B}),
\end{aligned}$$

and can be solved with efficient solvers. The update for $\boldsymbol{B}$ can be solved similarly.

The pseudo-solution for $\boldsymbol{C}$ is expressed as $\mathrm{vec}\{\boldsymbol{C}\} = (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^\dagger\boldsymbol{X}^\mathsf{T}\boldsymbol{z}$, with

$$\begin{aligned}
\boldsymbol{X} &= \boldsymbol{S}^\mathsf{T}(\boldsymbol{P}_2^\mathsf{T}\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1^\mathsf{T}\boldsymbol{P}_1)\boldsymbol{S} \boxtimes \boldsymbol{I}_K,\\
\boldsymbol{z} &= \mathrm{vec}\{(\boldsymbol{Y}_H^{(3)})^\mathsf{T}(\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S}\}.
\end{aligned}$$

Similarly, for $\widetilde{\boldsymbol{C}}_M$, we have

$$\boldsymbol{X} = \sqrt{\lambda}(\boldsymbol{S}^\mathsf{T}\boldsymbol{S} \boxtimes \boldsymbol{I}_{K_M}) \text{ and } \boldsymbol{z} = \sqrt{\lambda}\,\mathrm{vec}\{(\boldsymbol{Y}_M^{(3)})^\mathsf{T}\boldsymbol{S}\}.$$

Appendix B
Constrained updates in Algorithm 2

At each iteration of the ADMM scheme, we aim at solving the following equations:

$$(1+\rho)\boldsymbol{S}_r = \boldsymbol{A}_r\boldsymbol{B}_r^\mathsf{T} + \rho(\boldsymbol{Z}+\boldsymbol{U}), \tag{23}$$

$$\begin{aligned}
&\boldsymbol{C}(\boldsymbol{S}^\mathsf{T}(\boldsymbol{P}_2^\mathsf{T}\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1^\mathsf{T}\boldsymbol{P}_1)\boldsymbol{S} + \rho\boldsymbol{I}_R)\\
&= (\boldsymbol{Y}_H^{(3)})^\mathsf{T}(\boldsymbol{P}_2 \boxtimes \boldsymbol{P}_1)\boldsymbol{S} + \rho(\boldsymbol{Z}+\boldsymbol{U}),
\end{aligned} \tag{24}$$

$$\widetilde{\boldsymbol{C}}_M(\boldsymbol{S}^\mathsf{T}\boldsymbol{S} + \rho\boldsymbol{I}_R) = (\boldsymbol{Y}_M^{(3)})^\mathsf{T}\boldsymbol{S} + \rho(\boldsymbol{Z}+\boldsymbol{U}). \tag{25}$$

For each equation, $\boldsymbol{Z}$ is the projection of the considered variable onto the space of non-negative matrices, and $\boldsymbol{U}$ denotes the dual variable for each subproblem [38]. The scalar ρ controls the convergence speed of the algorithm and is chosen according to [39].

Below, we present the ADMM framework for solving (24): the updates for $\boldsymbol{S}$ and $\widetilde{\boldsymbol{C}}_M$ can be handled in a similar fashion.

Algorithm 4: ADMM update for (24)
input : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\boldsymbol{S}$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$; ρ, R, `iter`
output: $\boldsymbol{Z} \in \mathbb{R}_+^{K\times R}$
initialization: $\boldsymbol{Z} = \boldsymbol{U} = \boldsymbol{0}_{K\times M}$;
for $m \in \{1, \dots$ `iter`$\}$ **do**
  $\boldsymbol{C} \leftarrow$ Solve eq. (24) using normal equations;
  $\boldsymbol{Z} \leftarrow [\boldsymbol{C} - \boldsymbol{U}]_+$;
  $\boldsymbol{U} \leftarrow \boldsymbol{U} + \boldsymbol{Z} - \boldsymbol{C}$.
end

Here, the operator $[\cdot]_+$ zeroes out the negative values of the operand.

Appendix C
Spatial degradation matrices

Here, we explain in details how the degradation matrices are constructed. For this appendix, we consider that $\boldsymbol{P}_1 = \boldsymbol{P}_2$. As in previous works, $\boldsymbol{P}_1$ is constructed as $\boldsymbol{P}_1 = \boldsymbol{S}_1\boldsymbol{T}_1$, where $\boldsymbol{T}_1$ is a blurring matrix and $\boldsymbol{S}_1$ is a downsampling matrix.

The blurring matrix is constructed from a Gaussian blurring kernel $\boldsymbol{\phi} \in \mathbb{R}^{q\times 1}$ (in our case, $q = 9$) with a standard deviation $\sigma = \frac{q\sqrt{2\log 2}}{4}$. For $m \in \{1, \dots, q\}$ and $m' = m - \lceil \frac{q}{2} \rceil$, we have

$$\phi(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-m'^2}{2\sigma^2}\right).$$

Thus, $\boldsymbol{T}_1 \in \mathbb{R}^{I\times I}$ can be expressed as

$$\boldsymbol{T}_1 = \begin{bmatrix}
\phi(\lceil \frac{q}{2} \rceil) & \dots & \phi(q) & 0 & \dots & 0 \\
\vdots & \ddots & & \ddots & \ddots & \vdots \\
\phi(1) & & \ddots & & \ddots & 0 \\
0 & \ddots & & \ddots & & \phi(q) \\
\vdots & \ddots & \ddots & & \ddots & \vdots \\
0 & \dots & 0 & \phi(1) & \dots & \phi(\lceil \frac{q}{2} \rceil)
\end{bmatrix}.$$

The downsampling matrix $\boldsymbol{S}_1 \in \mathbb{R}^{I_H\times I}$, with downsampling ratio d, is made of I_H independant rows such that for $i \in \{1, \dots, I_H\}$, $(\boldsymbol{S}_1)_{i,2+(i-1)d} = 1$ and the other coefficients are zeros.

References

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[2] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *2nd IEEE Workshop on Hyperspectral Image and Signal Process.: Evolution in Remote Sens.*, 2010, pp. 1–4.

[3] L. Nus, *Méthodes rapides de traitement d'images hyperspectrales. Application à la caractérisation en temps réel du matériau bois*, PhD Thesis, University of Lorraine, France, 2019.

[4] G. A. Shaw and H. K. Burke, "Spectral imaging for remote sensing," *Lincoln laboratory journal*, vol. 14, no. 1, pp. 3–28, 2003.

[5] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 29–56, 2017.

[6] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012.

[7] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, 2015.

[8] M. Simoes, J. M. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015.

[9] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Multi-band image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, 2016.

[10] Q. Li, W.-K. Ma, and Q. Wu, "Hyperspectral super-resolution: Exact recovery in polynomial time," in *2018 IEEE SSP*. IEEE, 2018, pp. 378–382.

[11] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *Advances in neural information processing systems*, 2004, pp. 1141–1148.

[12] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Computational intelligence and neuroscience*, vol. 2008, 2008.

[13] H. Liu, R. Wu, and W.-K. Ma, "Is there any recovery guarantee with coupled structured matrix factorization for hyperspectral super-resolution?," in *2019 IEEE CAMSAP*. IEEE, 2019, pp. 480–484.

[14] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.

[15] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral Super-Resolution: Combining Low Rank Tensor and Matrix Structure," in *2018 IEEE ICIP*, Oct. 2018, pp. 3318–3322.

[16] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral Super-Resolution with Coupled Tucker Approximation: Identifiability and SVD-based algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 931–946, 2020.

[17] G. Zhang, X. Fu, K. Huang, and J. Wang, "Hyperspectral super-resolution: A coupled nonnegative block-term tensor decomposition approach," in *2019 IEEE CAMSAP*, 2019, event-place: Guadeloupe, West Indies.

[18] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *arXiv e-prints*, p. arXiv:2006.10248, June 2020.

[19] A. Eckardt, J. Horack, F. Lehmann, D. Krutz, J. Drescher, M. Whorton, and M. Soutullo, "DESIS (DLR earth sensing imaging spectrometer for the ISS-muses platform)," in *2015 IEEE IGARSS*, 2015, pp. 1457–1459.

[20] H. Kaufmann, K. Segl, S. Chabrillat, S. Hofer, T. Stuffler, A. Mueller, R. Richter, G. Schreier, R. Haydn, and H. Bach, "EnMAP a hyperspectral sensor for environmental mapping and analysis," in *2006 IEEE IGARSS*, 2006, pp. 1617–1619.

[21] T. Hilker, M. A. Wulder, N. C. Coops, J. Linke, G. McDermid, J. G. Masek, F. Gao, and J. C. White, "A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on landsat and MODIS," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1613–1627, 2009.

[22] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. M. Van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sensing of Environment*, vol. 133, pp. 193–209, 2013.

[23] R. A. Borsoi, T. Imbiriba, J. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, and C. Zare, A. .and Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review," *arXiv preprint arXiv:2001.07307*, 2020.

[24] R. A. Borsoi, T. Imbiriba, and J. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 116–127, 2020.

[25] T. Imbiriba, R. A. Borsoi, and J. M. Bermudez, "Generalized linear mixing model accounting for endmember variability," in *2018 IEEE ICASSP*, 2018, pp. 1862–1866.

[26] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. M. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE J. Sel. Topics Signal Process.*, 2021.

[27] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[28] P. Comon, "Tensors: A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, 2014.

[29] L. De Lathauwer, "Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.

[30] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. and Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[31] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sensing of Environment*, vol. 115, no. 7, pp. 1603–1616, 2011.

[32] A. Zare and K. C. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, 2013.

[33] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.

[34] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[35] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, 2008.

[36] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, 2008.

[37] V. Simoncini, "Computational methods for linear matrix equations," *SIAM Review*, vol. 58, no. 3, pp. 377–441, 2016.

[38] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011.

[39] K. Huang, N. D. Sidiropoulos, and A. P. Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5052–5065, 2016.

[40] G. Boutry, M. Elad, G. H. Golub, and P. Milanfar, "The generalized eigenvalue problem for nonsquare pencils using a minimal perturbation approach," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 2, pp. 582–601, 2005.

[41] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments-an estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, 2004.

[42] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab 3.0*, Available online, 2016.

[43] R. E. Roger and J. F. Arnold, "Reliably estimating the noise in AVIRIS hyperspectral images," *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1951–1962, 1996.

[44] *Sentinel-2 Document Library*.

[45] *LANDSAT Spectral response*.

[46] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogrammetric Eng. and Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

[47] J.M.P. Nascimento and J.M.B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 43, no. 4, pp. 898–910, 2005.

[48] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical als algorithms for nonnegative matrix factorization," *Neural computation*, vol. 24, no. 4, pp. 1085–1105, 2012.

[49] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2306–2320, 2015.

[50] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Process. Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[51] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 211–224, 2013.