MultiHU-TD: Multi-feature Hyperspectral Unmixing Based on Tensor Decomposition

Mohamad Jouni, Mauro Dalla Mura, Lucas Drumetz, Pierre Comon

MultiHU-TD: Multi-feature Hyperspectral Unmixing Based on Tensor Decomposition

Mohamad Jouni, *Member, IEEE,* Mauro Dalla Mura, *Senior Member, IEEE,* Lucas Drumetz, *Member, IEEE,* and Pierre Comon, *Fellow, IEEE*

***Abstract*—Hyperspectral unmixing allows to represent mixed pixels as a set of pure materials weighted by their abundances. Spectral features alone are often insufficient, so it is common to rely on other features of the scene. Matrix models become insufficient when the hyperspectral image is represented as a high-order tensor with additional features in a multimodal, multi-feature framework. Tensor models such as Canonical polyadic decomposition allow for this kind of unmixing, but lack a general framework and interpretability of the results. In this paper, we propose an interpretable methodological framework for low-rank Multi-feature hyperspectral unmixing based on tensor decomposition (MultiHU-TD) which incorporates the abundance sum-to-one constraint in the Alternating optimization ADMM algorithm, and provide in-depth mathematical, physical and graphical interpretation and connections with the extended linear mixing model. As additional features, we propose to incorporate mathematical morphology and reframe a previous work on neighborhood patches within MultiHU-TD. Experiments on real hyperspectral images showcase the interpretability of the model and the analysis of the results. `Python` and `MATLAB` implementations are made available on GitHub.**

***Index Terms*—Interpretability, Tensor decomposition, Hyperspectral unmixing, Extended linear mixing model, Blind source separation.**

I. Introduction

HYPERSPECTRAL IMAGING refers to the acquisition of images of a scene over a wide and almost continuous spectrum. A hyperspectral image (HSI) contains pixels that can cover areas of pure or mixed materials and amounts to a high spectral feature diversity [1], [2]. These characteristics allow to perform blind source separation (BSS) [3]–[5] on the observed spectral signatures to blindly extract those of pure materials (*sources*), also called *endmembers (EMs)*, and their per-pixel (*per-sample*) abundances. This case of BSS is known as hyperspectral unmixing (HU), which is an active research topic with several applications like remote sensing, chemometrics, biomedical imagery, etc [6]–[15]. HU allows to understand and quantify the physical components of a scene.

A significant part of research in BSS and HU relies on matrix factorization with additional constraints that aim at modeling the context of the problem. Consequently, an observed data matrix $\boldsymbol{M} \in \mathbb{R}^{I \times J}$ (i.e., with I pixel samples and J spectral features) is decomposed into two factor matrices $\boldsymbol{A} \in \mathbb{R}^{I \times R}$ and $\boldsymbol{B} \in \mathbb{R}^{J \times R}$ such that:

$$\boldsymbol{M} = \boldsymbol{A}\boldsymbol{B}^{\mathsf{T}} = \sum_{r=1}^{R} \boldsymbol{a}_r \boldsymbol{b}_r^{\mathsf{T}} \tag{1}$$

where R is the number of latent components to be estimated, and $\boldsymbol{a}_r$ and $\boldsymbol{b}_r$ are the columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, $\forall\, r \in \{1, \ldots, R\}$. As such, the columns of $\boldsymbol{B}$ represent the estimated source signals, and the rows of $\boldsymbol{A}$ represent the per-sample abundances of the sources. The decomposition is often carried out by minimizing the generic cost function [16]:

$$\underset{\boldsymbol{A},\boldsymbol{B}}{\operatorname{argmin}} \|\boldsymbol{M} - \boldsymbol{A}\boldsymbol{B}^{\mathsf{T}}\|_F^2 + r(\boldsymbol{A}) + r(\boldsymbol{B}) \tag{2}$$

where $r(\cdot)$ encodes the imposed constraints and/or regularizations to enforce desirable properties on the solutions.

In the case of HU, a classical approach is the nonnegative matrix factorization (NMF), which relies on the linear mixing model (LMM) of the observed HSI matrix (see Fig. 1). Hence, $\boldsymbol{A}$ and $\boldsymbol{B}$ are element-wise *nonnegative*, which applies also in most domains of BSS (other than HU) where the interpretability of the factor matrices is important. Moreover, the rows of $\boldsymbol{A}$ are subject to the *abundance sum-to-one constraint (ASC)*, which means that each row sums to 1:

$$\sum_{r=1}^{R} a_{ir} = 1 \quad \forall\, i \in \{1, \ldots, I\} \tag{3}$$

which applies to domains where the coefficients of the decomposition are proportions.

When only few materials concur in the mixture for each pixel, sparsity is imposed on the abundances [17]. Finally, real HSIs often contain spectral variabilities (SVs) in the sources, e.g., variations in the EMs due to local physico-chemical variations, illumination changes or topographic effects. In order to account to these SVs, the extended linear mixing model (ELMM) was proposed to extend the LMM to account to said SVs, which is an active topic that has seen a lot of progress recently [11]–[14].

A. Tensor Analysis of HSIs

An HSI can be treated as a data cube [18]–[23] (i.e., a third-order tensor with two spatial and one spectral dimensions). However, sometimes the HSI does not come alone but is associated with additional modalities such as:

This work is partly supported by grant ANR FuMultiSPOC (ANR-20-ASTR-0006), and partly by Région Auvergne-Rhône-Alpes grant "Pack Ambition International 2021" (21-007356-01FONC, 21-007356-02INV).

M. Jouni, M. Dalla Mura, and P. Comon are with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: mohamad.jouni@grenoble-inp.fr; mauro.dalla-mura@grenoble-inp.fr; pierre.comon@grenoble-inp.fr). M. Dalla Mura is also with Institut Universitaire de France (IUF), France.

L. Drumetz is with IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Brest, France (e-mail: lucas.drumetz@imt-atlantique.fr).

Manuscript received ...; revised ....

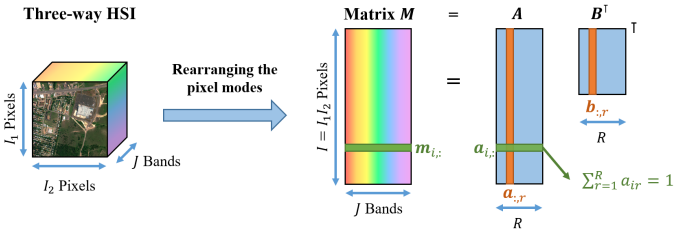


Fig. 1: Classical matrix-based HU using NMF (i.e., LMM)

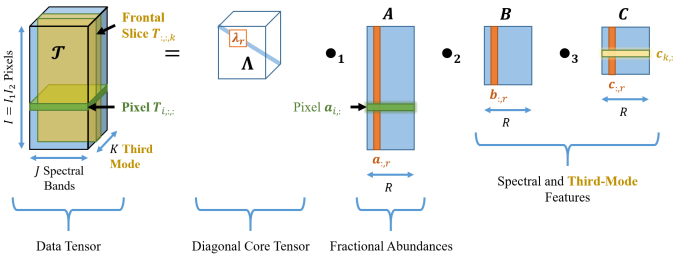


Fig. 2: Canonical polyadic decomposition of a third-order tensor. The tensor is formed of K matricized HSIs that are stacked along the third mode. A frontal slice $\boldsymbol{\mathcal{T}}_{:,:,k}$ (in yellow) represents a matricized HSI and is associated with one row of $\boldsymbol{C}$. A horizontal (pixel) slice $\boldsymbol{\mathcal{T}}_{i,:,:}$ (in green) represents a matrix of features and is associated with one row of $\boldsymbol{A}$.

- A time series or multi-angular data of HSI images [24].
- The HSI is combined with images acquired by different sensors (e.g., panchromatic, multispectral and LiDAR fusion) [25]–[28].
- Some spatial features are extracted from the HSI (such as in spectral-spatial classification problems [29]–[32]).

Such scenarios have recently also concerned other areas of BSS such as multi-channel signal processing [33]–[35] and multidimensional biomedical signal and image processing [36]–[40]. In the aforementioned scenarios, the data are represented natively as *tensors*[1] [41], and the challenge usually boils down to the proper modeling of a joint factorization of multivariate representations without losing the multimodal structure, and hence its interpretation in terms of BSS.

Among these scenarios, we focus on the case of HU where the HSI is associated with an additional set of features in the form of a new tensor mode, which we coin as *Multi-feature HU based on Tensor Decomposition* (MultiHU-TD). For example, we consider a set of spatial features extracted from the image itself that can be considered as new modes. Moreover, the pixels are rearranged in lexicographic order, promoting *low-rank* tensor decomposition[2]. In this scenario, there are some challenging questions to answer such as: how can we jointly perform a constrained factorization in such settings? And how can we interpret the extracted factors? An analysis of the literature shows that there are works that perform NMF with additional constraints [42]–[44], and others that consider the case of multimodal inputs with coupled NMF [45], [46], but this is different from considering data as tensors in our case.

As the native structure of our data is a tensor, we consider the problem in terms of tensor decomposition [41], which is the natural framework for processing multimodal data in the signal and image processing community [47]–[49]. There are many types of decomposition, such as tucker decomposition, block term decomposition (BTD), canonical polyadic decomposition (CPD), etc [50]. However, we choose CPD thanks to the diagonality of its core tensor which allows the interaction only between similarly indexed columns in the factor matrices, which *naturally* promotes *source separation* and a straightforward *interpretability* of the extracted components in unmixing, both of which are core aspects of our work.

In the third-order case, as illustrated in Fig. 2, CPD decomposes a data tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I\times J\times K}$ into a diagonal core tensor $\boldsymbol{\Lambda} \in \mathbb{R}^{R\times R\times R}$ and 3 factor matrices $\{\boldsymbol{A} \in \mathbb{R}^{I\times R}, \boldsymbol{B} \in \mathbb{R}^{J\times R}, \boldsymbol{C} \in \mathbb{R}^{K\times R}\}$, each representing one of the 3 modes of $\boldsymbol{\mathcal{T}}$ respectively, such that:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\Lambda} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{C} \tag{4}$$

where $\bullet_d$ denotes the mode-d product (product along the d-th mode), further described in expressions (6), (13b), and (13c).

CPD extends NMF to high-order data and can adopt all of its features, especially that of imposing constraints. CPD is often computed by minimizing the cost function [41]:

$$\underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}}{\operatorname{argmin}} \|\boldsymbol{\mathcal{T}} - \boldsymbol{\Lambda} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{C}\|_F^2 + r(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}) \tag{5}$$

where $r(\cdot)$ encodes the imposed constraints. Note that the nonnegative constraint ensures the existence of a minimum; in fact, without an appropriate regularisation term $r(\cdot)$, the above cost function could admit only an infimum, which may not be reachable [51]. On the other hand, with an appropriate regularisation, Problem (5) is well posed.

B. Related Works and Limitations

In the context of MultiHU-TD, CPD has been used with multitemporal/angular HSIs [24] as well as with HSIs having an additional diversity of extracted neighborhood patches [14] (see Fig. 7 for a 5×5 patch-HSI tensor). Moreover, some works [29]–[31] jointly considered HSIs with spatial features extracted by mathematical morphology (MM) filters [52]–[55] in the framework of multi-feature scene classification (see Fig. 8). These works show that CPD is a suitable approach for joint decomposition. However, they present some limitations.

1) Algorithmic perspective: The nonnegative constraint is implemented in [56] by projection onto the nonnegative orthant, which exhibits some computational issues [48]. In [14], nonnegative alternating least squares (ALS) is used where ASC is also naively implemented by projecting the abundances on the unit simplex, contrary to the common practice in the matrix case [17] where ASC is embedded in the updates. In [24], the nonnegative CPD is computed using the projected

[1]A tensor can be represented as a multidimensional array. The order of a tensor refers to the number of its array's indices, which is also the number of its *modes*. For example, a tensor of dimensions $I \times J \times K$ is said to have three *modes*, and is called a *third-order tensor*. Data sets with order 3 or above are described as *high-order* tensors.

[2]Pixels form only one mode, even if images are often seen as 2D objects. This suppresses the high-rankness introduced by the complex spatial features of the whole scene [15], which is an inconvenience for BSS methods [26].

compressed ALS (ProCo-ALS) algorithm, which is considerably fast [47] but not so flexible with additional constraints. Finally, in [29]–[31], an alternative algorithm is proposed based on alternating optimization alternating direction method of multipliers (AO-ADMM) [49] with compression and nonnegative constraints, which is flexible and stable with large datasets, but has not yet addressed MultiHU-TD which requires further modeling (i.e., sparsity, ASC).

2) Interpretability: The work of [24] faced a certain challenge in interpreting the third-mode factors, perhaps due to the naive employment of CPD. In [14], a link was established between CPD and ELMM but was not deeply investigated as it was restricted to the case of patches and tested only with synthetic data. Moreover, it faced another challenge in interpreting the factors, which poses an ambiguity on its performance and the meaning of the extracted features. Finally, [14] and [29]–[31] perform tensor decomposition with spatial features. However, the used spatial features can allow limited flexibility (e.g., patches), or the interpretation of the factor matrices was not addressed, noting here that incorporating them with CPD showed improvement in supervised classification (e.g., MM).

In this paper, we wish to consider such operations in a BSS framework from the lens of MultiHU-TD with in-depth interpretability. This presents us with two main challenges:

- Tuning AO-ADMM to incorporate ASC, which is challenging due to the multilinear structure of CPD, particularly in modeling the samples as a convex combination of the spectral sources in a multimodal setting.
- Exploring the meaning of the extracted features in these conditions.

C. Contributions

To our knowledge, imposing ASC in CPD or AO-ADMM as a natural extension of NMF in [17] has not been done. Furthermore, we are interested in finding a generalized framework for MultiHU-TD favouring the interpretation of its results under any third-mode diversity. More precisely, our contributions to jointly deal with these limitations are the following:

- We propose a methodological framework for dealing with MultiHU-TD based on AO-ADMM by Huang [49], and expand it to incorporate ASC with joint nonnegativity and sparsity. The proposed AO-ADMM-ASC is a general algorithm that can be applied in other domains of BSS where convex combinations of sources apply.
- We establish a unified framework for the interpretability of MultiHU-TD. In particular, the link between ELMM and CPD [14] is expanded by providing in-depth physical and graphical insights for better interpretability of the CPD model and its factors.
- We propose to include MM as spatial features to perform a spectral-spatial HU and demonstrate the aforementioned points. We also revise [14] and provide detailed interpretations on the cases of patches and MM, which has not been addressed in any of the previous works [14], [29]–[31]. This analysis also shows that MM is better suited since it embeds physically meaningful features (scale and brightness of objects) into HU unlike patches.

That said, note that our main goal is to reason about the interpretability of this factorization and to describe this framework rather than to propose yet another HU algorithm.

The remainder is organized as follows. In section III, we introduce some background. In section IV, we detail the proposed framework. In section V, we present our experiments and results. Finally, we draw out some conclusions in VI.

II. Notations and Definitions

Table I shows a list of notations for the different types of objects used throughout the paper: scalars, vectors, matrices, tensors, and array dimensions and indices. Table II denotes the types of observed data in the paper, their dimensions and different ways of indexing.

Type	Font style	Example
Scalars	unformatted lowercase	a, b, c, t
Vectors	bold lowercase	$\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{t}$
Matrices	bold uppercase	$\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{T}$
Tensors	bold calligraphic	$\boldsymbol{\mathcal{T}}$
Dimension	unformatted uppercase	I, J, K, R
Indices	lowercase version of the spanned dimension	i, j, k, r

TABLE I: Array notations

Table III denotes the different ways to slice and unfold a third-order tensor. The *mode unfolding* (or *matricization*) of a tensor means to reshape it into a matrix by fixing the targeted mode and rearrange the others in lexicographic order.

Table IV denotes the factor matrices of an NMF (matrix case) or CPD (tensor case). Mode-1, Mode-2 and Mode-3 correspond to the modes of *pixels*, spectral *bands*, and set of extracted spatial features (*transforms*) respectively.

We use the notation "$\mathrm{diag}\{\boldsymbol{v}\}$" to refer to the diagonal matrix whose entries are the elements of any vector $\boldsymbol{v}$.

The *outer product* of two vectors $\boldsymbol{a} \in \mathbb{R}^I$ and $\boldsymbol{b} \in \mathbb{R}^J$ results in a matrix $\boldsymbol{M} \in \mathbb{R}^{I \times J}$ as follows:

$$\boldsymbol{M} = \boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^{\mathsf{T}} \Longleftrightarrow m_{i,j} = a_i b_j$$
$$\forall\, i \in \{1, \ldots, I\},\ \ \forall\, j \in \{1, \ldots, J\}$$

The *outer product* of three vectors $\boldsymbol{a} \in \mathbb{R}^I$, $\boldsymbol{b} \in \mathbb{R}^J$ and $\boldsymbol{c} \in \mathbb{R}^K$ results in a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I \times J \times K}$ as follows:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} \Longleftrightarrow t_{i,j,k} = a_i b_j c_k$$
$$\forall\, i \in \{1, \ldots, I\},\ \ \forall\, j \in \{1, \ldots, J\},\ \ \forall\, k \in \{1, \ldots, K\}$$

The *mode-d product* $\bullet_d$ represents the product of a tensor by a matrix along the d-th mode. For example, assuming that we have $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{L \times M \times N}$, $\boldsymbol{A} \in \mathbb{R}^{I \times L}$ and $\boldsymbol{B} \in \mathbb{R}^{J \times M}$, the mode-1 and mode-2 product of $\boldsymbol{\mathcal{G}}$ by $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively results in a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I \times J \times N}$ defined as:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{G}} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \Longleftrightarrow t_{ijn} = \sum_{l=1}^{L} \sum_{m=1}^{M} G_{lmn}\, a_{il} b_{jm} \quad (6)$$

III. Background

In this section, we briefly review the existing notions in the literature upon which we base our algorithm and generalized interpretation of the MultiHU-TD framework. First, we

Type	Symbol	Dimensions (pixel × band × transform)	i-th pixel	j-th band	k-th transform	(i,j,k)-th element
HSI matrix	$\boldsymbol{M}$	$I \times J$	$\boldsymbol{m}_{i,:}$	$\boldsymbol{m}_{:,j}$	-	$m_{i,j}$
HSI tensor	$\boldsymbol{\mathcal{T}}$	$I \times J \times K$	$\boldsymbol{T}_{i,:,:}$	$\boldsymbol{T}_{:,j,:}$	$\boldsymbol{T}_{:,:,k}$	$t_{i,j,k}$

TABLE II: The pixels are rearranged in *lexicographic order* spanning the first mode, so I is the total number of pixels. The symbol ":" in the index indicates a span of the whole mode. For example, $\boldsymbol{m}_{i,:}$ and $\boldsymbol{m}_{:,j}$ represent the i-th row and j-th column *vectors* of $\boldsymbol{M}$ respectively (see Fig. 1), and $\boldsymbol{T}_{:,:,k}$ represents the k-th frontal *matrix* slice of $\boldsymbol{\mathcal{T}}$ (see Fig. 2).

Variable	Symbol	Dimensions
Horizontal slice	$\boldsymbol{T}_{i,:,:}$	$J \times K$
Lateral slice	$\boldsymbol{T}_{:,j,:}$	$I \times K$
Frontal slice	$\boldsymbol{T}_{:,:,k}$	$I \times J$
Mode-1 unfolding	$\boldsymbol{T}_{(1)}$	$JK \times I$
Mode-2 unfolding	$\boldsymbol{T}_{(2)}$	$IK \times J$
Mode-3 unfolding	$\boldsymbol{T}_{(3)}$	$IJ \times K$

TABLE III: Tensor slicing and mode-unfolding.

Factor Mat.	Symbol	Dimensions	Row index	Col. index	Element index
Mode-1	$\boldsymbol{A}$	$I \times R$	$\boldsymbol{a}_{i,:}$	$\boldsymbol{a}_{:,r}$	$a_{i,r}$
Mode-2	$\boldsymbol{B}$	$J \times R$	$\boldsymbol{b}_{j,:}$	$\boldsymbol{b}_{:,r}$	$b_{j,r}$
Mode-3	$\boldsymbol{C}$	$K \times R$	$\boldsymbol{c}_{k,:}$	$\boldsymbol{c}_{:,r}$	$c_{k,r}$

TABLE IV: The factor matrices, each corresponding to one of the matrix or tensor modes. R is the number of rank-1 additive terms in the decomposition.

explain how ASC is applied in NMF [17] as the proposed framework extends this for tensor decomposition. Then, we give a brief account on ELMM, including graphical and visual interpretations, which will be the basis for the proposed interpretation. Finally, we discuss the link between CPD and ELMM preliminarily presented in [14].

A. NMF with ASC, Nonnegativity, and Sparsity

In the NMF case [17], when sparsity (ℓ_1 norm) and ASC are imposed on the abundances, (1) becomes:

$$\begin{aligned} &\underset{\boldsymbol{A},\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{M}-\boldsymbol{A}\boldsymbol{B}^{\mathsf{T}}\|_F^2+\alpha\|\boldsymbol{A}\|_1 \\ &\text{s.t. } \boldsymbol{A}\succeq 0,\ \boldsymbol{B}\succeq 0,\ \sum_{r=1}^{R} a_{i,r}=1\ |_{\forall i\in\{1,\dots,I\}} \end{aligned} \tag{7}$$

where $\alpha > 0$, and $\succeq$ denotes element-wise nonnegativity. A simple strategy to embed ASC goes by stacking a row vector in $\boldsymbol{B}$ and a column vector in $\boldsymbol{M}$ such that [17]:

$$\tilde{\boldsymbol{M}} = \left[\boldsymbol{M} \mid \delta \mathbf{1}_{I\times 1}\right],\ \tilde{\boldsymbol{B}} = \begin{bmatrix}\boldsymbol{B}\\ \delta \mathbf{1}_{1\times R}\end{bmatrix}, \tag{8}$$

where δ is a constant that is usually set as the mean of $\boldsymbol{M}$, and the last row of $\tilde{\boldsymbol{B}}$ is reset to δ after each iteration. This operation ensures that ASC is softly embedded in NMF since $\forall i \in \{1,\dots,I\}$ we have:

$$m_{i,J+1} = \sum_{r=1}^{R} a_{i,r} b_{J+1,r} = \sum_{r=1}^{R} a_{i,r}\delta = \delta \tag{9}$$

corresponding to $\sum_{r=1}^{R} a_{i,r} = 1$. Then, (7) becomes:

$$\underset{\boldsymbol{A},\tilde{\boldsymbol{B}}}{\operatorname{argmin}} \frac{1}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{A}\tilde{\boldsymbol{B}}^{\mathsf{T}}\|_F^2+\alpha\|\boldsymbol{A}\|_1 \text{ s.t. } \boldsymbol{A}\succeq 0, \tilde{\boldsymbol{B}}\succeq 0 \tag{10}$$

There are many algorithms proposed in the literature that deal with sparse NMF and ASC, which are out of the scope of this work [42], [43]. In our case, we extend NMF within the AO-ADMM framework for CPD. NMF then becomes a special case for order-2 tensors. This tensor extension, proposed in Section IV-A is referred to as AO-ADMM-ASC.

B. ELMM

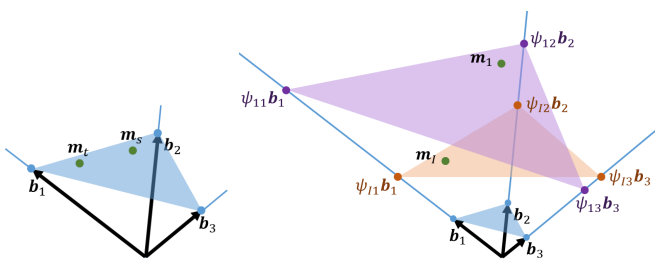
While LMM is seen as a direct approach for HU, it cannot model SVs represented by nonlinear effects or illumination conditions. One way to account to said effects is through ELMM [12], which in general assumes additional degrees of freedom that account to said SVs *at the pixel level* by introducing a *pixel-dependent* SV function $\boldsymbol{f}_i : \mathbb{R}^J \to \mathbb{R}^J$ $\forall\, i \in \{1,\dots,I\}$, which maps each EM $\boldsymbol{b}_{:,r}$ $\forall\, r \in \{1,\dots,R\}$ to a new spectral signature $\boldsymbol{b}_{:,r}^{(i)}$ that best reflects the targeted SVs:

$$\boldsymbol{m}_i = \sum_{r=1}^{R} a_{ir}\boldsymbol{f}_i(\boldsymbol{b}_{:,r}) = \sum_{r=1}^{R} a_{ir}\boldsymbol{b}_{:,r}^{(i)}. \tag{11}$$

For example, in the case of different illumination conditions, this can be represented as a scaling factor for each pixel on the EMs. In the following, we present the parts that are at the basis of the interpretability of our proposed framework.

When ASC is imposed in LMM, all the pixels will lie on the convex hull of the set of estimated EMs (i.e., the columns of $\boldsymbol{B}$[3]), and the fractional abundances in each row of $\boldsymbol{A}$ define the coordinates of each pixel on the convex hull, which is illustrated in Fig. 3a. With the introduction of ELMM, the pixels will not lie on the same simplex anymore as each pixel is mapped to a new set of EMs, which is illustrated in Fig. 3b

[3]Here, we note that unless the EMs are not affinely independent, which is unlikely, and $R \leq J+1$, then the convex hull is a simplex.



(a) LMM (b) ELMM (scalar factors)

Fig. 3: Graphical comparison between LMM and ELMM (scalar factors) in the case of three spectral signatures $\{\boldsymbol{b}_{:,1}, \boldsymbol{b}_{:,2}, \boldsymbol{b}_{:,3}\}$ and two example pixels $\{\boldsymbol{m}_{s,:}, \boldsymbol{m}_{t,:}\}$. The relative coordinates in the simplices are conserved.

in the case where the SV is modeled by a scaling factor ψ_{ir} that is pixel- and EM-dependent such that $\boldsymbol{b}_r^{(i)} = \psi_{ir}\boldsymbol{b}_{:,r}$ [12]. Accordingly, the new spectral signatures $\boldsymbol{b}_r^{(i)}$ are allowed to move only along the directions of $\boldsymbol{b}_{:,r}$.

In the case where the SVs are modeled with scaling factors, we denote by $\boldsymbol{\Psi} \in \mathbb{R}^{I\times R}$ the matrix whose entries are the scaling factors $\psi_{ir}\ \forall\, i \in \{1,\dots,I\}$ and $\forall\, r \in \{1,\dots,R\}$. Then, the following are equivalent:

$$\boldsymbol{M} = (\boldsymbol{A} \boxdot \boldsymbol{\Psi})\boldsymbol{B}^{\mathsf{T}} \iff \tag{12a}$$

$$\boldsymbol{m}_{i,:} = \sum_{r=1}^{R} a_{ir}\psi_{ir}\boldsymbol{b}_{:,r}^{\mathsf{T}} = \boldsymbol{a}_{i,:}\,\mathrm{diag}\{\boldsymbol{\Psi}_{i,:}\}\boldsymbol{B}^{\mathsf{T}} = \boldsymbol{a}_{i,:}\boldsymbol{\Psi}_{(i)}\boldsymbol{B}^{\mathsf{T}} \tag{12b}$$

where $\boxdot$ is the Hadamard product. $\boldsymbol{\Psi}_{i,:} \in \mathbb{R}^{R}$ is the i-th row of $\boldsymbol{\Psi}$, and $\boldsymbol{\Psi}_{(i)} \in \mathbb{R}^{R\times R}$ represents the diagonal matrix formed from $\boldsymbol{\Psi}_{i,:}$. These mathematical, graphical, and visual relationships are the key to elaborate and interpret the case of CPD in Section IV-B.

C. CPD and ELMM

In the work of [14], a representation of CPD in terms of ELMM has been presented, which will be reported here. CPD decomposes a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I\times J\times K}$ such that:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\Lambda} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{C} \Leftrightarrow \boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \lambda_{r,r,r}\ \boldsymbol{a}_{:,r} \otimes \boldsymbol{b}_{:,r} \otimes \boldsymbol{c}_{:,r} \tag{13a}$$

$$\Leftrightarrow \boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \boldsymbol{a}_{:,r} \otimes \boldsymbol{b}_{:,r} \otimes \boldsymbol{\psi}_{:,r} \tag{13b}$$

$$\Leftrightarrow t_{i,j,k} = \sum_{r=1}^{R} a_{i,r}\ b_{j,r}\ \psi_{k,r} \tag{13c}$$

where $t_{i,j,k}$ is a (scalar) entry of $\boldsymbol{\mathcal{T}}$, R is the number of estimated sources, and $\boldsymbol{\Lambda} \in \mathbb{R}^{R\times R\times R}$ is a diagonal tensor that absorbs the ℓ_2-norms of the columns of the factor matrices [41], [48]. In (13b) and (13c), we suppress the expression of $\boldsymbol{\Lambda}$ by absorbing its entries in the columns of $\boldsymbol{C}$, resulting in $\boldsymbol{\Psi} \in \mathbb{R}^{K\times R}$ whose columns are the scaled version of those of $\boldsymbol{C}$ such that $\boldsymbol{\psi}_{:,r} = \lambda_{r,r,r}\boldsymbol{c}_{:,r}\ \forall\, r$.

As shown in [14], given that $\boldsymbol{T}_{:,:,k}$ denotes the k-th frontal slice of $\boldsymbol{\mathcal{T}}$, $\forall\, k \in \{1,\dots,K\}$, and assuming that $\boldsymbol{\Psi}_{(k)} \in \mathbb{R}^{R\times R}$ represents the diagonal matrix formed from the row $\boldsymbol{\psi}_{k,:} \in \mathbb{R}^{R}$ of the factor $\boldsymbol{\Psi}$, one can write:

$$\boldsymbol{T}_{:,:,k} = \boldsymbol{A}\,\mathrm{diag}\{\boldsymbol{\psi}_{k,:}\}\boldsymbol{B}^{\mathsf{T}} = \boldsymbol{A}\boldsymbol{\Psi}_{(k)}\boldsymbol{B}^{\mathsf{T}} = \boldsymbol{A}\tilde{\boldsymbol{f}}_k(\boldsymbol{B})^{\mathsf{T}} \tag{14a}$$

$$\iff \boldsymbol{t}_{i,:,k} = \sum_{r=1}^{R} a_{i,r}\ (\boldsymbol{b}_{:,r}\psi_{k,r}) = \sum_{r=1}^{R} a_{i,r}\boldsymbol{f}_k(\boldsymbol{b}_{:,r}) \tag{14b}$$

where $\boldsymbol{t}_{i,:,k}$ is the i-th pixel row of $\boldsymbol{T}_{:,:,k}$. From the perspective of each frontal slice, CPD resembles a regularized coupled NMF of the frontal slices where $\boldsymbol{A}$ is a common factor:

$$\underset{\boldsymbol{A},\boldsymbol{B}^{(k)}}{\mathrm{argmin}} \|\boldsymbol{T}_{:,:,k} - \boldsymbol{A}\boldsymbol{B}^{(k)\mathsf{T}}\|_F^2 \text{ s.t. } \boldsymbol{A} \succeq 0, \boldsymbol{B}^{(k)} \succeq 0 \tag{15}$$

where $\boldsymbol{B}^{(k)} = \tilde{\boldsymbol{f}}_k(\boldsymbol{B})|_{\forall\, k\in\{1,\dots,K\}}$, which has an analogous expression to the case of ELMM. The latter was used in [14] as a way to impose a spatial smoothing (15) on the abundances within a patch of neighboring pixels. This link between CPD and ELMM was preliminary presented and restricted to an application of patches with only simulated data, and the meaning of the SVs was not explored.

In Section IV-B, we propose a generalized in-depth interpretation of MultiHU-TD, and in Section IV-C we present how to incorporate spatial features (e.g., patches and MM) in such a setting.

IV. Proposed MultiHU-TD Framework

In this section, we present the proposed MultiHU-TD framework based on CPD. First, we talk about the implementation of AO-ADMM-ASC. Then, we provide an interpretation of MultiHU-TD including the physical meaning of ASC and the ELMM model in the tensor case. Finally, we propose to include spatial features as examples of the third modality.

A. AO-ADMM-ASC with Nonnegativity and Sparsity

In CPD, after imposing nonnegativity on the factor matrices, and sparsity and ASC on the abundances, (5) becomes:

$$\begin{aligned} &\underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}}{\mathrm{argmin}} \|\boldsymbol{\mathcal{T}} - \boldsymbol{\Lambda} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{C}\|_F^2 + \alpha\|\boldsymbol{A}\|_1 \\ &\text{s.t. } \boldsymbol{A} \succeq 0,\ \boldsymbol{B} \succeq 0,\ \boldsymbol{C} \succeq 0,\ \sum_{r=1}^{R} a_{i,r} = 1\ |_{\forall i\in\{1,\dots,I\}} \end{aligned} \tag{16}$$

In (16), since $\boldsymbol{A}$ adopts the sum-to-one constraint on its rows, it is hence enough to normalize only the columns of $\boldsymbol{B}$ and $\boldsymbol{C}$. In principle, these ℓ_2-norms are absorbed in $\boldsymbol{\Lambda}$, but for the sake of consistency, we use the variable $\boldsymbol{\Psi} = \boldsymbol{\Lambda}\boldsymbol{C}$ instead of $\boldsymbol{\Lambda}$ and $\boldsymbol{C}$, as explained in Section III-C. In order to solve (16), we propose an algorithm inspired by AO-ADMM [49], where the factor matrices are updated in an alternating way and where each update of a factor matrix is optimized as an ADMM subproblem.

1) ASC Solution: In order to model the problem as close as possible to LMM, we need a tensor decomposition algorithm embedding both the ASC and the non-negativity of factors as constraints. To this end, the strategy we follow is to extend the concept from NMF to CPD by stacking:

- a row vector to $\boldsymbol{B}$ (i.e., $\boldsymbol{b}_{J+1,:} \in \mathbb{R}^{R}$)
- a lateral slice to $\boldsymbol{\mathcal{T}}$ (i.e., $\boldsymbol{T}_{:,J+1,:} \in \mathbb{R}^{I\times 1\times K}$)

such that $\sum_{r=1}^{R} a_{i,r} = 1|_{\forall i\in\{1,\dots,I\}}$ is ensured.

In general, $\boldsymbol{T}_{:,J+1,:}$ can be constructed such that $\forall i \in \{1,\dots,I\}$ and $\forall k \in \{1,\dots,K\}$:

$$t_{i,J+1,k} = \sum_{r=1}^{R} a_{i,r}b_{J+1,r}\psi_{k,r} \tag{17}$$

So if we set:

- $\boldsymbol{T}_{:,J+1,K} = \delta\boldsymbol{1}_I$, i.e., $t_{i,J+1,K} = \delta\ \forall i \in \{1,\dots,I\}$
- $b_{J+1,r} = \delta\psi_{K,r}^{-1}\ \ \forall r \in \{1,\dots,R\}$,

where δ is the mean of $\boldsymbol{\mathcal{T}}$, by substituting the expressions in (17) for $k = K$, we have:

$$t_{i,J+1,K} = \sum_{r=1}^{R} a_{i,r} b_{J+1,r} \psi_{K,r} \tag{18a}$$

$$\implies \delta = \delta \sum_{r=1}^{R} a_{i,r} \psi_{K,r}^{-1} \psi_{K,r} = \delta \sum_{r=1}^{R} a_{i,r} \tag{18b}$$

which implies that $\sum_{r=1}^{R} a_{i,r} = 1$.

At the end of each AO-ADMM iteration, $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{B}$ have to be updated, which boils down to a matrix and a vector updates after each iteration. We denote by $\tilde{\boldsymbol{\mathcal{T}}}$ and $\tilde{\boldsymbol{B}}$ the extensions of $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{B}$ with the additional lateral slice and row vector respectively, roughly described as follows:

$$\tilde{\boldsymbol{\mathcal{T}}} = \left[\boldsymbol{\mathcal{T}} \mid \boldsymbol{T}_{:,J+1,:}\right], \; \tilde{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{b}_{J+1,:} \end{bmatrix}, \tag{19}$$

Then, expression (16) becomes:

$$\underset{\boldsymbol{A},\tilde{\boldsymbol{B}},\boldsymbol{\Psi}}{\operatorname{argmin}} \|\tilde{\boldsymbol{\mathcal{T}}} - \boldsymbol{\mathcal{I}} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \tilde{\boldsymbol{B}} \underset{3}{\bullet} \boldsymbol{\Psi}\|_F^2 + \alpha \|\boldsymbol{A}\|_1 \quad \text{s.t. } \boldsymbol{A} \succeq 0, \; \tilde{\boldsymbol{B}} \succeq 0, \; \boldsymbol{\Psi} \succeq 0 \tag{20}$$

where $\boldsymbol{\mathcal{I}}$ is a diagonal tensor of ones.

2) ADMM Updates: At this stage, solving (20) with AO-ADMM becomes simple. We demonstrate the ADMM sub-problem updates for each factor matrix starting with $\boldsymbol{A}$.

Supposing that $\tilde{\boldsymbol{T}}_{(1)}$ represents the mode-1 unfolding of $\tilde{\boldsymbol{\mathcal{T}}}$, we can write the sub-problem of $\boldsymbol{A}$ as follows:

$$\boldsymbol{A} = \underset{\boldsymbol{A}}{\operatorname{argmin}} \frac{1}{2} \|\tilde{\boldsymbol{T}}_{(1)} - \tilde{\boldsymbol{W}}_{(A)} \boldsymbol{A}^{\mathsf{T}}\|_F^2 + \alpha \|\boldsymbol{A}\|_1 \quad \text{s.t. } \boldsymbol{A} \succeq 0 \tag{21}$$

where $\tilde{\boldsymbol{W}}_{(A)} \in \mathbb{R}^{(J+1)K \times I} = \tilde{\boldsymbol{B}} \odot \boldsymbol{\Psi}$ represents the Khatri-Rao product [41]. By introducing the splitting variable $\bar{\boldsymbol{A}} = \boldsymbol{A}^{\mathsf{T}}$, expression (21) becomes:

$$\underset{\boldsymbol{A},\bar{\boldsymbol{A}}}{\operatorname{argmin}} \frac{1}{2} \|\tilde{\boldsymbol{T}}_{(1)} - \tilde{\boldsymbol{W}}_{(A)} \bar{\boldsymbol{A}}\|_F^2 + \alpha \|\boldsymbol{A}\|_1 \quad \text{s.t. } \bar{\boldsymbol{A}} = \boldsymbol{A}^{\mathsf{T}} \text{ and } \boldsymbol{A} \succeq 0 \tag{22}$$

Adopting ADMM for (22), the updates of $\bar{\boldsymbol{A}}$ and $\boldsymbol{A}$ become:

$$\begin{aligned} \bar{\boldsymbol{A}} &\leftarrow (\tilde{\boldsymbol{W}}_{(A)}^{\mathsf{T}} \tilde{\boldsymbol{W}}_{(A)} + \rho \boldsymbol{I})^{-1} (\tilde{\boldsymbol{W}}_{(A)}^{\mathsf{T}} \tilde{\boldsymbol{T}}_{(1)} + \rho (\boldsymbol{A} + \boldsymbol{U}_{(A)})^{\mathsf{T}}) \\ \boldsymbol{A} &\leftarrow \max(0, \bar{\boldsymbol{A}}^{\mathsf{T}} - \boldsymbol{U}_{(A)} - \frac{\alpha}{\rho}) \\ \boldsymbol{U}_{(A)} &\leftarrow \boldsymbol{U}_{(A)} + \boldsymbol{A} - \bar{\boldsymbol{A}}^{\mathsf{T}} \end{aligned} \tag{23}$$

where $\boldsymbol{U}_{(A)} \in \mathbb{R}^{I \times R}$ is called the dual variable.

Similarly, the updates of $\tilde{\boldsymbol{B}}$ and $\boldsymbol{\Psi}$ become:

$$\begin{aligned} \bar{\boldsymbol{B}} &\leftarrow (\tilde{\boldsymbol{W}}_{(B)}^{\mathsf{T}} \tilde{\boldsymbol{W}}_{(B)} + \rho \boldsymbol{I})^{-1} (\tilde{\boldsymbol{W}}_{(B)}^{\mathsf{T}} \tilde{\boldsymbol{T}}_{(2)} + \rho (\boldsymbol{B} + \boldsymbol{U}_{(B)})^{\mathsf{T}}) \\ \tilde{\boldsymbol{B}} &\leftarrow \max(0, \bar{\boldsymbol{B}}^{\mathsf{T}} - \boldsymbol{U}_{(B)}) \\ \boldsymbol{U}_{(B)} &\leftarrow \boldsymbol{U}_{(B)} + \tilde{\boldsymbol{B}} - \bar{\boldsymbol{B}}^{\mathsf{T}} \end{aligned} \tag{24}$$

$$\begin{aligned} \bar{\boldsymbol{\Psi}} &\leftarrow (\tilde{\boldsymbol{W}}_{(\Psi)}^{\mathsf{T}} \tilde{\boldsymbol{W}}_{(\Psi)} + \rho \boldsymbol{I})^{-1} (\tilde{\boldsymbol{W}}_{(\Psi)}^{\mathsf{T}} \tilde{\boldsymbol{T}}_{(3)} + \rho (\boldsymbol{\Psi} + \boldsymbol{U}_{(\Psi)})^{\mathsf{T}}) \\ \boldsymbol{\Psi} &\leftarrow \max(0, \bar{\boldsymbol{\Psi}}^{\mathsf{T}} - \boldsymbol{U}_{(\Psi)}) \\ \boldsymbol{U}_{(\Psi)} &\leftarrow \boldsymbol{U}_{(\Psi)} + \boldsymbol{\Psi} - \bar{\boldsymbol{\Psi}}^{\mathsf{T}} \end{aligned} \tag{25}$$

where $\tilde{\boldsymbol{T}}_{(2)}$ and $\tilde{\boldsymbol{T}}_{(3)}$ are the mode-2 and mode-3 unfoldings of $\tilde{\boldsymbol{\mathcal{T}}}$, $\tilde{\boldsymbol{W}}_{(B)} = \boldsymbol{A} \odot \boldsymbol{\Psi}$ and $\tilde{\boldsymbol{W}}_{(\Psi)} = \boldsymbol{A} \odot \tilde{\boldsymbol{B}}$ are the Khatri-Rao products, and $\boldsymbol{U}_{(B)}$ and $\boldsymbol{U}_{(\Psi)}$ are the dual variables.

Finally, for order-2 tensors, this model becomes equivalent to solving NMF (10). The implementation of AO-ADMM-ASC is summarized in Algorithm 1. The code is available on GitHub in `Python`[4] and `MATLAB`[5].

Algorithm 1 AO-ADMM-ASC

Require: $\boldsymbol{\mathcal{T}}$, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Psi}$, $\boldsymbol{U}_{(A)}, \boldsymbol{U}_{(B)}, \boldsymbol{U}_{(\Psi)}$, α
  Initialize $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Psi}$;
  Initialize $\boldsymbol{U}_{(A)}, \boldsymbol{U}_{(B)}, \boldsymbol{U}_{(\Psi)}$ to zero;
  repeat
    Set $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{\mathcal{T}}}$ based on (19);
    Set $\boldsymbol{B} \leftarrow \tilde{\boldsymbol{B}}$ and $\boldsymbol{\mathcal{T}} \leftarrow \tilde{\boldsymbol{\mathcal{T}}}$;
    for $\forall \boldsymbol{D} \in \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Psi}\}$ **do**
      $\tilde{\boldsymbol{W}}_{(D)} = \odot_{\boldsymbol{J} \neq \boldsymbol{D}} \boldsymbol{J}$;
      $\rho = \text{trace}(\tilde{\boldsymbol{W}}_{(D)}^{\mathsf{T}} \tilde{\boldsymbol{W}}_{(D)})/R$; [49]
      Update $\boldsymbol{D}$ with either (23), (24), or (25);
    end for
    Absorb the column-wise ℓ_2-norms of $\boldsymbol{B}$ into $\boldsymbol{\Psi}$ such that $\boldsymbol{\psi}_{:,r} \leftarrow \boldsymbol{\psi}_{:,r} \; \|\boldsymbol{b}_{:,r}\|_2 \quad \forall r \in \{1, \ldots, R\}$
    Normalize the columns of $\boldsymbol{B}$
  until Termination criterion (e.g., number of iterations)
  return $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Psi}$

B. Interpreting Tensor-based unmixing, ASC, and ELMM

Here, we build upon what has been presented in Sections III-B and III-C as methodological, physical, and graphical bases for the MultiHU-TD interpretation. For that, we first draw the analogies between the expressions of Sections III and IV by starting from the interpretation of the matrix case and elaborating that of the tensor case. In the process, we break down the physical meaning of ASC and that of the so-called SV function of ELMM in MultiHU-TD at the base of the composition of the frontal slices of the tensor. Then, we visualize the expressions in order to interpret MultiHU-TD through graphical representations of subspaces while commenting on the physical role of the extracted factors and the number of latent components R.

1) Interpretation of ASC: In the matrix case, assuming a matricized HSI $\boldsymbol{M}$ (that is, after reordering the two pixel modes into one mode in lexicographic order) such that:

$$\boldsymbol{M} = \boldsymbol{A}\,\boldsymbol{B}^{\mathsf{T}} \tag{26}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ represent the estimated abundances and endmembers respectively, the physical meaning of ASC is that it constrains the columns of $\boldsymbol{B}$ to form a simplex. Then, the rows of $\boldsymbol{A}$ (which sum to one) represent the position of the pixels on said simplex. This is visualized in Fig. 3.

In the tensor case, we assume a tensor $\boldsymbol{\mathcal{T}}$ whose CPD is expressed as:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{I}} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{\Psi} \tag{27}$$

[4] https://github.com/mhmdjouni/MultiHU-TD-Python
[5] https://github.com/mhmdjouni/MultiHU-TD-MATLAB

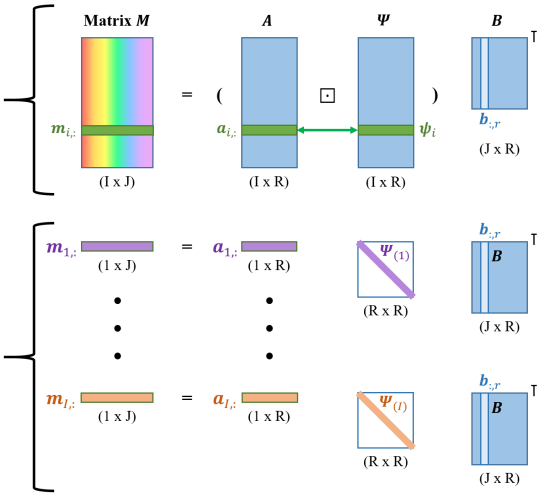


Fig. 4: Our visualization of equations (12a) (Hadamard product) and (12b) (matrix product). The color code of the bottom part follows that of Fig. 3b. We have $\boldsymbol{\Psi}_{(i)} = \text{diag}\{\boldsymbol{\Psi}_{i,:}\}$.

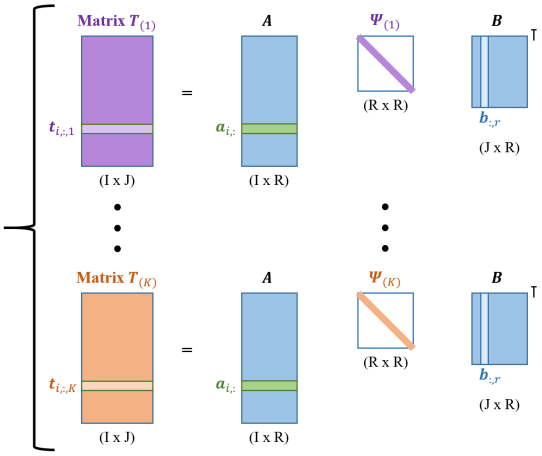
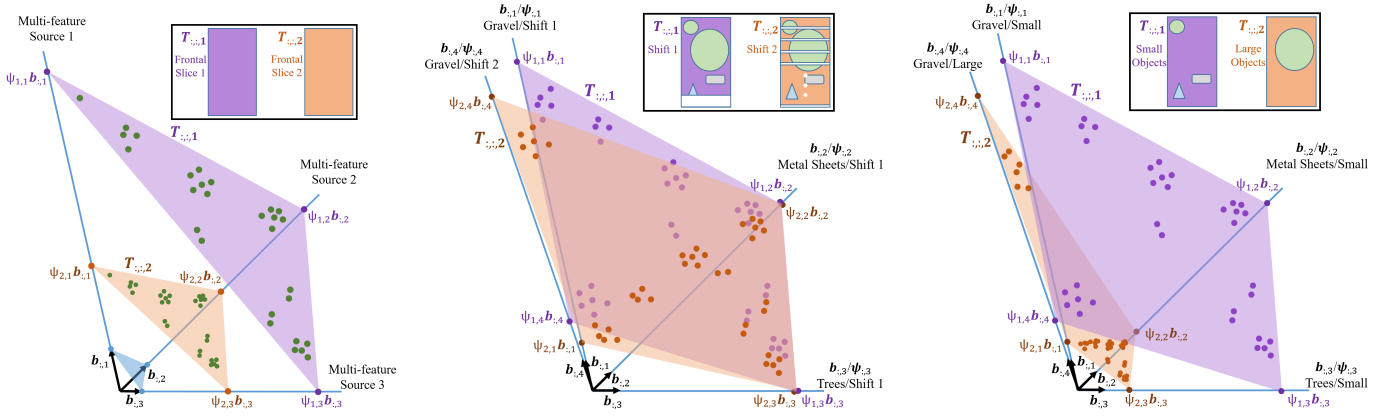


Fig. 5: Our visualization of equation (14a). The color code is made analogous to that of Fig. 2 and follows that of Fig. 6a. We have $\boldsymbol{\Psi}_{(k)} = \text{diag}\{\boldsymbol{\psi}_{k,:}\}$.



(a) ELMM-CPD (General) (b) ELMM-CPD (Patches) (c) ELMM-CPD (MM)

Fig. 6: Graphical representations of (a) CPD with $R = 3$ components in the case of three spectral signatures $\{\boldsymbol{b}_{:,1}, \boldsymbol{b}_{:,2}, \boldsymbol{b}_{:,3}\}$ and two frontal slices $\{\boldsymbol{T}_{:,:,1}, \boldsymbol{T}_{:,:,K}\}$, and (b) patch-CPD and (c) MM-CPD with $R = 4$ components in the case of four spectral signatures $\{(\boldsymbol{b}_{:,1}, \boldsymbol{\psi}_{:,1}), \ldots, (\boldsymbol{b}_{:,4}, \boldsymbol{\psi}_{:,4})\}$ and two frontal slices $\{\boldsymbol{T}_{:,:,1}, \boldsymbol{T}_{:,:,2}\}$. Very correlated vectors can be seen as "spectral bundles" with different third-modality characteristics. The relative coordinates of the pixels in the convex hulls must be the same since $\boldsymbol{A}$ is common for all the frontal slices.

where $\boldsymbol{\mathcal{I}}$ is a diagonal tensor of ones, $\boldsymbol{A}$ and $\boldsymbol{B}$ represent the estimated abundances and endmembers respectively, and $\boldsymbol{\Psi}$ represents the factor matrix of the third modality.

Here, each slice of the tensor $\boldsymbol{T}_{:,:,k}$ $\forall k \in \{1, \ldots, K\}$ represents a matricized HSI similar to $\boldsymbol{M}$ (e.g., corresponding to acquisitions at different dates in a time series, at different angles in multi-angular acquisitions or at different scales in a multi-scale decomposition), and we have the following linear relationship:

$$\boldsymbol{T}_{:,:,k} = \boldsymbol{A}\,(\text{diag}\{\Psi_{k,:}\}\,\boldsymbol{B}^{\mathsf{T}}) = \boldsymbol{A}\,\boldsymbol{B}^{(k)\,\mathsf{T}}\,\forall k, \tag{28}$$

where $\Psi_{k,:}$ represents the k-th row of $\boldsymbol{\Psi}$. First, let us look at the properties of the aforementioned expression:

- If we look at each slice separately, the physical meaning of applying ASC is similar to that of the matrix case where the simplex is formed out of the columns of $\boldsymbol{B}^{(k)}$. In fact, we have that $\boldsymbol{B}^{(k)} = \boldsymbol{B}\ \text{diag}\{\Psi_{k,:}\}$, which means that the columns of $\boldsymbol{B}^{(k)}$ are only scaled versions of those of $\boldsymbol{B}$ such that:

$$\boldsymbol{b}^{(k)}_{:,r} = \psi_{k,r}\boldsymbol{b}_{:,r}\ \ \forall r \in \{1, \ldots, R\} \tag{29}$$

- If we consider all the slices together, we notice that the abundance matrix $\boldsymbol{A}$ is common to all of them. Moreover, we notice that the factor matrix of estimated endmembers $\boldsymbol{B}$ (obtained through CPD) is at the base of their estimated endmembers, influenced only by the corresponding scaling factors in $\Psi_{k,:}$, which encode the corresponding third-mode features.
- Given that $\boldsymbol{A}$ and $\boldsymbol{B}$ factorize the physical and data structures along the pixel and spectral mode respectively, they are indepedent of the third-mode differences in the hyperspectral scene between the slices. One could even construct a matrix $\boldsymbol{M}^{(\text{CPD})}$ from the first two factor

matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (obtained through CPD) such that:

$$\boldsymbol{M}^{(\text{CPD})} = \boldsymbol{A}\,\boldsymbol{B}^{\mathsf{T}} \tag{30}$$

where $\boldsymbol{B}$ is independent of the spectral variabilities present along the third mode.

- As $\boldsymbol{A}$ and $\boldsymbol{B}$ factorize the pixel and spectral information, the third-mode factor matrix $\boldsymbol{\Psi}$ encodes the changes between the slices along the third mode where the k-th row $\Psi_{k,:}$ is associated to the k-th slice. Intuitively speaking, this allows some degrees of freedom to express the nonlinearities along the third mode in a linear sense, which is reflected on the level of each slice by scaling the columns of $\boldsymbol{B}$ as expressed in (29).

Finally, assuming that we have K slices in the tensor, imposing ASC is the equivalent of having K simplices whose edges, defined by the columns of $\boldsymbol{B}^{(k)}$ $\forall k \in \{1, \ldots, K\}$, can move only along the directions of the columns of $\boldsymbol{B}$ such that:

- The positions of the edges of each simplex is defined by the corresponding scaling factor $\psi_{k,r}$, which encodes the third-mode physical property (whether it is time, morphological properties such as scale and brightness, neighborhood pixels, etc) of the r-th estimated endmember $\boldsymbol{b}_{:,r}$ in the k-th tensor slice $\boldsymbol{T}_{:,:,k}$.
- For instance, an estimated endmember $\boldsymbol{b}_{:,r}$ can be relevant in a given slice $\boldsymbol{T}_{:,:,1}$ due to a high factor $\psi_{1,r}$, such that $\boldsymbol{b}^{(1)}_{:,r} = \psi_{1,r}\boldsymbol{b}_{:,r}$, but also have a low contribution in another slice $\boldsymbol{T}_{:,:,2}$ due to a low factor $\psi_{2,r}$, such that $\boldsymbol{b}^{(2)}_{:,r} = \psi_{2,r}\boldsymbol{b}_{:,r}$. We can see that $\boldsymbol{b}_{:,r}$ remains independent of the physical entity that the third mode represents, but also that $\boldsymbol{b}^{(1)}_{:,r}$ and $\boldsymbol{b}^{(2)}_{:,r}$ move along the direction of $\boldsymbol{b}_{:,r}$ (due to the scaling factors $\psi_{1,r}$ and $\psi_{2,r}$) based on the effect that the the third-mode physical entity applies on $\boldsymbol{b}_{:,r}$ in the given slice.
- Since $\boldsymbol{A}$ is common to all the slices, the positions of the pixels are relatively fixed to each of the K simplices.

This is demonstrated in Fig. 5 and 6 in the manuscript.

2) Interpretation of tensor-based ELMM: First, we note that expressions (14a) and (14b) are analogous to the ELMM expression (12b). The major difference between the two cases is that in CPD, the scaling factors are *frontal slice-dependent* ($\psi_{k,r}$), while in ELMM, they are pixel-dependent ($\psi_{i,r}$). Second, we visualize (12a) and (12b) in Fig. 4, and (14a) and (14b) in Fig. 5. Looking at (14a), the frontal slices $\boldsymbol{T}_{:,:,k}$ and the physical meaning that they represent have a direct influence on the SV function $\boldsymbol{f}_k$ and the interpretation of the SVs, which is simply reflected as scaling factors in each row of $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\psi}_{k,:}$ (or $\boldsymbol{\Psi}_{(k)} = \text{diag}\{\boldsymbol{\psi}_{k,:}\}$).

As a result, since the spatial and spectral information are factorized and represented by $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively, and since $\boldsymbol{A}$ and $\boldsymbol{B}$ are shared by all the frontal slices, then each frontal slice $\boldsymbol{T}_{:,:,k}$ is inherently differentiated through a set of R scaling factors $\{\psi_{k,1}, \ldots, \psi_{k,R}\}$. Consequently, the spectral information in each frontal slice $\boldsymbol{T}_{:,:,k}$ can be seen as the set of scaled sources $\{\boldsymbol{b}^{(k)}_{:,r} = \psi_{k,r}\boldsymbol{b}_{:,r}\}|_{\forall\, r\in\{1,\ldots,R\}}$, where $\{\boldsymbol{b}_{:,r}\}$ are the columns of $\boldsymbol{B}$ (independent of the slices), $\boldsymbol{b}^{(k)}_{:,r}$ are their spectral variations per frontal slice, and $\boldsymbol{\psi}_{k,:}$ encodes the scaling factors of these variations.

This also means that R is a major parameter that represents the degrees of freedom especially through the scaling factors of $\boldsymbol{\Psi}$, which then jointly encodes:

- the mode-3 evolution of the extracted components of $\boldsymbol{A}$ and $\boldsymbol{B}$ in its columns $\boldsymbol{\psi}_{:,r}$
- the per-slice modeling of the SVs in its rows $\boldsymbol{\psi}_{k,:}$.

Intuitively, when obtaining an augmented HSI tensor, one can say that the physical representations of any applied transformations (e.g., scale, illumination) [31] and any natural evolution of a scene (e.g., time series) [24], and resulting in $\boldsymbol{T}_{:,:,k}$, are reflected and observed through the matrix $\boldsymbol{\Psi}$ of the decomposition. Moreover, we point out the following:

- In CPD, there are as many SV functions (and simplices) as the frontal slices of the tensor, which is significantly lower than the number of pixels ($K \lll I$).
- In CPD, one row of R scaling factors in $\boldsymbol{\Psi}$ corresponds to a full frontal slice and is shared by all the pixel rows of $\boldsymbol{A}$, while in classical ELMM, each row of scaling factors in $\boldsymbol{\Psi}$ corresponds to one pixel of $\boldsymbol{M}$ and interacts with only one row of $\boldsymbol{A}$. This clearly appears when we compare (11) and (12b) to (14a) and (14b), and Fig. 4 to Fig. 5.
- On a graphical representation, when ASC is imposed, CPD suggests that each simplex contains I pixels as illustrated in Fig. 6a, such that the relative coordinates of the pixels inside each convex hull are the same since each row of $\boldsymbol{\Psi}$ interacts with all of $\boldsymbol{A}$.

In summary, having a third mode in HSI produces scaling factors in ELMM that absorb the SVs based on the physical meaning of the frontal slices along the third mode (e.g., time, patches, MM), which balances the extracted factors in $\boldsymbol{A}$ and $\boldsymbol{B}$ independently of said SVs. Moreover, the imposed value of R represents the number of extracted sources and scaling factors and has a major effect on the results and the SV interpretation. Since the extraction of materials is also driven by the third mode diversity, it is possible to expect a few more latent components than the number of pure "EMs" existing in the scene, with some spectral correlations as roughly demonstrated in Fig. 6. As R decreases, we tend towards having fewer degrees of freedom, where CPD tends towards extracting the EMs while applying a regularization on the observed pixels influenced by the physical meaning of the information across the third mode. As R increases, we tend towards having more degrees of freedom, where CPD tends towards extracting factors with multi-feature separability of the sources. We note that R should not be too high in order to avoid over-fitting[6] (given the relative low-rankness of the data) and to ensure uniqueness of the CPD [41], [51].

C. Examples of third-mode features: Spatial Features ¨(patches / mathematical morphology)

In this section, for the sake of comparison and illustration, we consider two examples of spatial features that augment a

[6]There is no exact value of the tensor rank, and finding a good estimate is an open challenge (as for matrix factorization problems), which is out of the scope of this paper. As a rule of thumb, R could be chosen by looking at the reconstruction error in the factorization, e.g. by choosing the smallest value of R providing an acceptable reconstruction, or by observing the elbow of the plot of singular values of the mode-1 unfolding of the data.

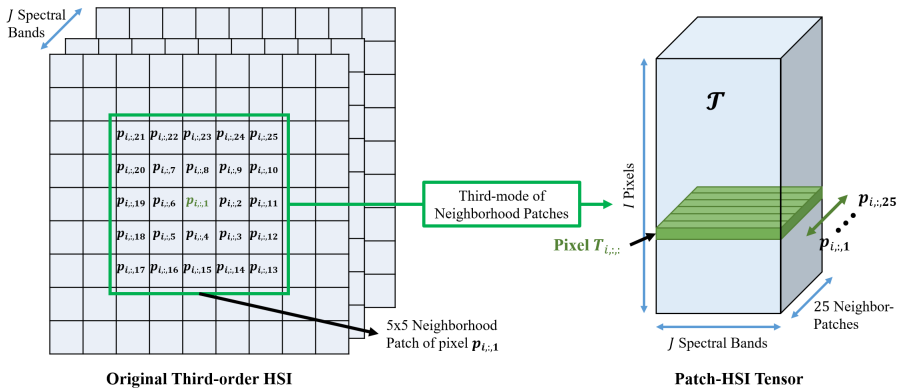


Fig. 7: An illustration of constructing a 5×5 Patch-HSI tensors based on [14].

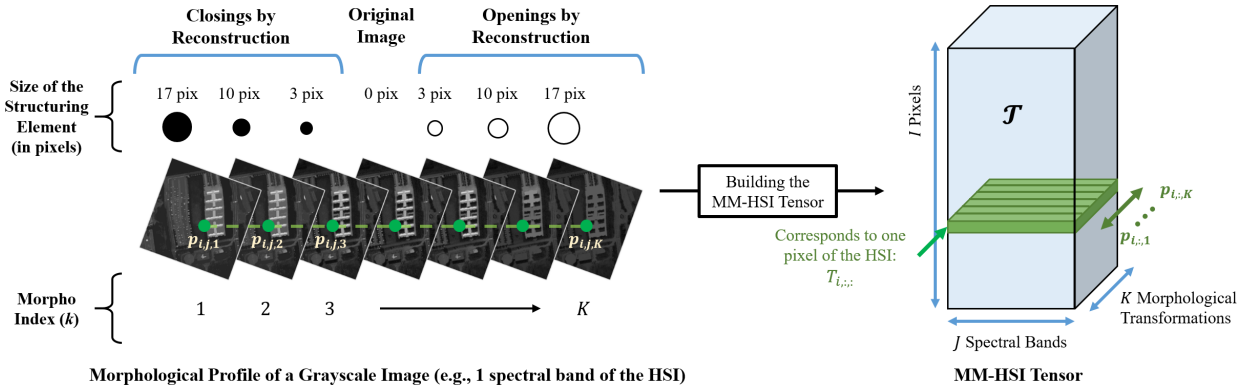


Fig. 8: Example of a sequential morphological filtering of a grayscale image (corresponding to one spectral band of the HSI of Pavia University) with Openings and Closings by Reconstruction using successive sizes of the structuring element, which is a disk in this case. The stacking of the transformations, with the original image corresponding to 0 size being placed in the middle, is referred to as the morphological profile of the image. The green spots mark the same pixel positioning in each of the transformations. Then, to create the MM-HSI tensor [31], the pixel positioning modes are reordered into lexicographic order along the first mode, and the yellow spots are stored along the third mode, in order to create the tensor.

HSI into a third-order tensor for MultiHU-TD: neighborhood patches and MM. Having these two types of features allow for a more comprehensive comparison in terms of the properties of the MultiHU-TD framework and its links to previous works on said features. We revisit the case of patches with additional insights, and introduce MM. Consequently, this helps demonstrate the interpretability of the model especially in terms of physical significance and the variation of R. We often refer to Fig. 5 and 6 for illustration.

1) Patches: We recall that the motivation for adding patches as features is to perform a spatial regularization by considering the spatial correlation of neighboring pixels [14]. Constructing a HSI tensor from neighborhood patches (coined as "Patch-HSI tensor") is illustrated in Fig. 7. In short, each pixel tube in the original HSI cube is taken with a patch of its neighboring pixels (of predetermined size), then the pixel and its neighbors are stacked as a horizontal slice $\boldsymbol{T}_{i,:,:}$ in the third-order tensor. That said, the first frontal slice $\boldsymbol{T}_{:,:,1}$ of said tensor is usually the matricized HSI (where the pixels represent one mode).

Here, we add that a Patch-HSI tensor has an inherently low-rank structure which is equal to that of the NMF of the matricized HSI, that is, the information contained along the third mode by each of the frontal slices of the tensor is almost essentially the same. In fact, the frontal slices are just *spatially-shifted versions* of the original image $\boldsymbol{T}_{:,:,1}$, and the values of these shifts correspond to a *small spatial kernel*, usually around 3×3 or 5×5. However, what is different in the tensor case is that when this *shifting* information is stacked along the third mode and CPD is imposed with such a low value of R, the model automatically applies an implicit smoothing of the pixels that belong to the same patch (i.e., the same horizontal slice of $\boldsymbol{\mathcal{T}}$). This is because the frontal slices are jointly factorized with the degrees of freedom of a single one of them, while also sharing the information of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Therefore, a main advantage over NMF is that one expects to extract the same sources with a patch-local smoothing of the SVs of the estimated EMs, where the SVs are balanced out in the form of scaling factors stored in the rows of $\boldsymbol{\Psi}$. An important note here is that the scaling factors stored in $\boldsymbol{\Psi}$ may not have a significant physical meaning.

Now, what happens when R increases? Since the information across the frontal slices are essentially the same (implying redundancy), the sources and abundances are expected to replicate, and we expect to observe slightly spatially-shifted versions of the abundance maps (i.e., in the columns of $\boldsymbol{A}$). In this case, the scaling factors in $\boldsymbol{\Psi}$ only indicate whether an

estimated EM in $\boldsymbol{B}$ corresponds to a certain spatial shifting or another. This point is roughly illustrated in Fig. 6b (inspired by Fig. 11) where we have three spectral sources: Gravel, Metal Sheets, and Trees, but CPD is carried out with $R = 4$. Here, the convex hull of $\boldsymbol{\mathcal{T}}_{:,:,1}$ gives a high scaling factor at $\boldsymbol{b}_1$ and a low factor at $\boldsymbol{b}_4$, while that of $\boldsymbol{\mathcal{T}}_{:,:,2}$ gives the opposite with almost the same quantity. This is due to the fact that the materials are present with almost the same quantity in both frontal slices. In other words, there may be a problem of redundancy if some components account for the same material with patches, which does happen in practice.

This problem does not occur when the third mode represents a physical meaning such as the case of MM.

2) Mathematical morphology: While using patches is efficient, it still ignores the physical properties of connected pixels, and the SVs are regularized indifferently among pixels belonging to different types of materials. On the other hand, morphological features [57], [58] take into account physical properties such as scale and brightness of objects and promote dealing with SV among pixels sharing these properties. Constructing a HSI tensor using MM (coined as "MM-HSI tensor") is illustrated in Fig. 8. In short, the matricized versions of the original HSI and the results of its morphological transformations are stacked as the frontal slices of the tensor[7].

Through MM, we emphasize the role of incorporating spatial diversities that add physical significance to the objects of the scene. As such, one expects that a MM-HSI tensor has a more complicated structure than that of a Patch-HSI tensor since its frontal slices contain additional context on the materials, such as their sizes and brightness levels. As such, imposing a low R promotes spectral smoothing of the SVs based on a morphological regularization of the abundances, while imposing a sufficiently high R promotes a distinctive spectral-morphological multi-feature separation of the materials, unlike Patch-HSI tensors.

Since each frontal slice is seen as a characteristic of spatial scale (i.e., size of objects in the scene) and / or brightness (which is particularly relevant for ELMM because scaling factors can be directly linked to brightness), then the scaling factors represented by $\boldsymbol{\Psi}$ indicate the quantitative correspondence of an extracted material to the aforementioned physical properties per frontal slice. This point is roughly illustrated in Fig. 6c (inspired by Fig. 12) where $\boldsymbol{\mathcal{T}}_{:,:,1}$ and $\boldsymbol{\mathcal{T}}_{:,:,2}$ characterize small and large objects respectively. Here, the convex hull of $\boldsymbol{\mathcal{T}}_{:,:,1}$ gives high scaling factors for $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ (corresponding to small objects), and a low scaling factor at $\boldsymbol{b}_4$ (corresponds to large objects), while that of $\boldsymbol{\mathcal{T}}_{:,:,2}$ gives the opposite.

V. Experiments and Results

In this section, we discuss the experiments and results of MultiHU-TD on real HSIs in terms of AO-ADMM-ASC (compared to Naive ASC [14]), extracted factors, ELMM interpretability, and qualitative comparisons between Patch-HSI and MM-HSI tensors with low and high values of the number of latent components. In each experiment, among 30 random initializations of the factor matrices, the result with the minimum root mean squared error (RMSE) is chosen based on (31a). The estimated EMs of $\boldsymbol{B}$ are identified based on their minimum spectral angular distance (SAD), in degrees, with respect to the reference EMs based on (31b).

$$\text{RMSE}\left(\boldsymbol{\mathcal{T}}, \hat{\boldsymbol{\mathcal{T}}}\right) = \frac{\|\boldsymbol{\mathcal{T}} - \hat{\boldsymbol{\mathcal{T}}}\|_F^2}{\|\boldsymbol{\mathcal{T}}\|_F^2} \tag{31a}$$

$$\text{SAD}(\boldsymbol{e}, \boldsymbol{b}) = \arccos\left(\frac{\boldsymbol{e} \cdot \boldsymbol{b}}{\|\boldsymbol{e}\|_2 \cdot \|\boldsymbol{b}\|_2}\right) \tag{31b}$$

The maps and plots shown in the experiments represent the *columns* of the factor matrices. Above each abundance map (AM), we show the material that corresponds to it with its minimum SAD value. We recall that a set of similarly indexed columns, e.g., $\{\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{\Psi}_1\}$, represent the abundance, spectral source signal, and third-mode source pattern (e.g., morphological print, shifting print) of one extracted material respectively. Since $\boldsymbol{\Psi}$ plays a crucial role in the interpretability of ELMM and MultiHU-TD, we also highlight the relevance of its row components. Each fixed index k in the plot corresponds to a row of $\boldsymbol{\Psi}$ and thus to a frontal slice in $\boldsymbol{\mathcal{T}}$, and the vertical grouping of points at said index, as indicated in Fig. 11c, 12c, 13c, and 14c, represents the scalars in that row. This also means that when projected on Fig. 6, in Patch-CPD, $k = 1$ contains the SV scaling factors that balance the convex hull of the original HSI and are responsible for its reconstruction, while in MM-CPD, the middle index does that. This will eventually show how MM accounts to physical effects in the scene, while patches do not.

That said, we note that we carried part of the experiments using sparse NMF with ASC [17] for the sake of qualitative comparison of the abundance maps and spectral sources obtained from the original HSI (i.e., the HSI matrix without additional filtering). Due to the difference in the type of information contained between the matrix and tensor cases, and since NMF does not apply in the framework of MultiHU-TD, these results will serve only as a reference for the extracted components of CPD as they do not serve the main aim and message of this work. For that reason, we include them in Appendix B with further reasoning and explanation of the NMF case analysis and its relevance to this work.

In each case analysis, we look into the components of $\boldsymbol{A}$ and $\boldsymbol{B}$ first, which visually and spectrally identify the materials, then we explain their correspondence to those of $\boldsymbol{\Psi}$, where we are interested in the significance of the third-mode patterns then their relevance to the original HSI. We note that quantitative validation of the AMs and EMs is usually not evident, especially in the case of the Urban HSI in Fig. 10b where the spatial ground-truth (GT) is not a real GT but actually just a reference, and is not usable for quantitative comparison. Moreover, there is neither a quantitative nor a qualitative reference for third-mode patterns in the literature, so highlight an in-depth qualitative analysis.

We want to consider HSIs which show objects with spatial features of different scale and brightness levels, for which urban areas are good candidates. For that, we choose two

[7]The details of constructing MM-HSI tensors are out of the scope of this paper, but can be found in Section 3 of [31].

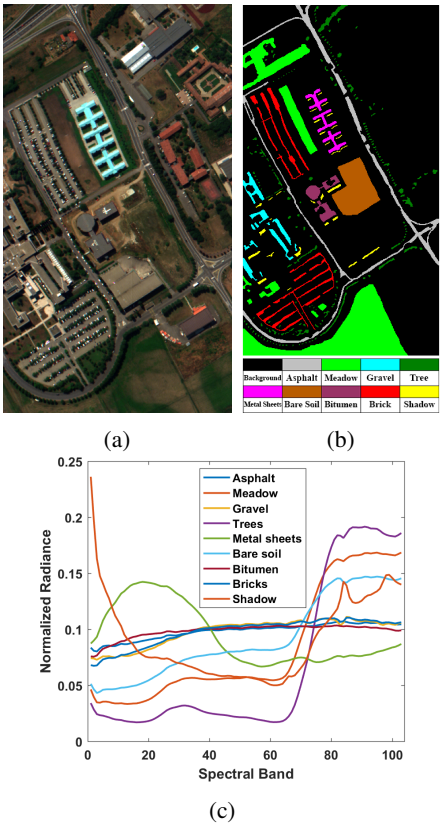


Fig. 9: 9a) Pavia in false colors. 9b) Pavia's spatial reference. 9c) Pavia's spectral reference extracted by averaging each class of the spatial reference.

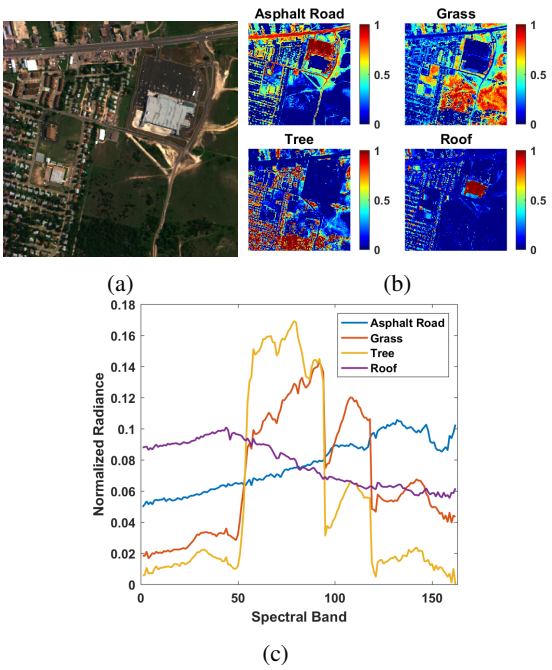


Fig. 10: 10a) Urban in false colors. 10b) Urban's spatial reference. 10c) Urban's spectral reference.

real HSIs[8]: *Pavia University* and *Urban*, shown in Fig. 9 and 10 respectively in false colors with their spatial and spectral references. It is worth noting for the sake of interpretation that the MM-HSI tensors are built following the extended morphological profile (EMP) technique used in [31] with openings by reconstruction (ObRs) and closings by reconstruction (CbRs), which correspond to bright and dark objects respectively, with varying sizes of the structuring element (SE), which in turns corresponds to the scales of objects. In the following experiments, 4 sizes of the SE are used, corresponding to 8 ObR and CbR transformations, then the dimension of the third mode is $K = 9$ where the original image (corresponding to scale 0) is placed in the middle as roughly shown in Fig. 8. The Patch-HSI tensors are built following [14] as shown in Fig. 7 with 3×3 patches. This means that the dimension of the third mode is $K = 3 \times 3 = 9$ where the original image is placed at $k = 1$. In addition to the real HSIs, we demonstrate the performance of the proposed framework under different levels of noise through a synthetic HSI tensor in Appendix C. We run our experiments with Intel® Core™ i7-1185G7, 32GB RAM 3200MHz LPDDR4.

A. Results Discussion - Pavia University

In this section, we present the experiments of the HSI of Pavia, but first, we note that the reference for this dataset is originally composed of training and testing sets, where each pixel is manually associated with one of 9 labeled classes as seen in Fig. 9b. The spectral reference in Fig. 9c is extracted by averaging the spectral signatures of each subset of pixels belonging to one class. In Fig. 9c, some classes have very similar spectral signatures, so, in the following, sometimes we refer to *Trees* and *Meadows* as *vegetation*, and to *Asphalt*, *Bitumen*, *Gravel*, and *Bricks* as *roads* or *roofs*, while *Bare Soil* may belong to either of both groups.

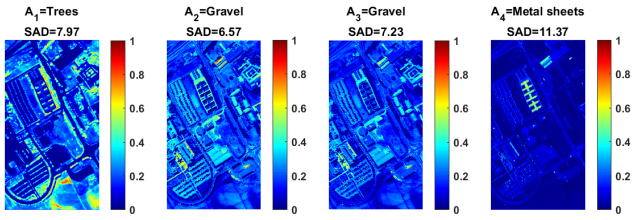
We start by comparing AO-ADMM-ASC and Naive ASC. After that, we focus on CPD and the ELMM analysis of the factors while interpreting the cases of patches and MM. For the MM-HSI tensor, our SEs are disks with the successive radii: $\{2, 7, 12, 17\}$ pixels. Both Patch- and MM-HSI tensors then have $K = 9$ frontal slices and dimensions $207400 \times 103 \times 9$. Finally, we find that $R = 4$ and $R = 8$ are the best for low and high values of the number of latent components respectively.

1) AO-ADMM-ASC: Here, we compare the RMSE results of MM-CPD between AO-ADMM-ASC and Naive ASC [14]. The results are shown in Table V, where we see that with AO-ADMM-ASC we gain in RMSE, which corresponds to a better estimation of the factors with respect to the observed tensor with a small difference in the execution time.

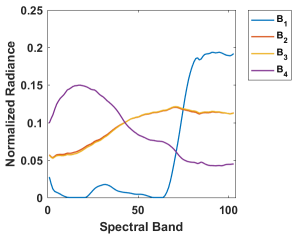
Algorithm	R	RMSE %	Time (s)
Naive ASC [14]	8	7.07	231
AO-ADMM-ASC	8	6.34	384

TABLE V: Pavia. The results of AO-ADMM-ASC and Naive ASC in terms of RMSE and execution time: R indicates the number of latent components. The results of the minimum RMSE are shown.

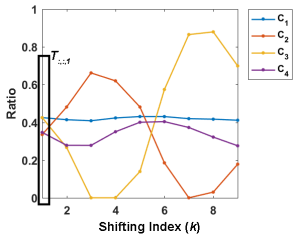
[8]The data sets with detailed information are available on the website: http://lesun.weebly.com/hyperspectral-data-set.html



(a) Components of $\boldsymbol{A}$

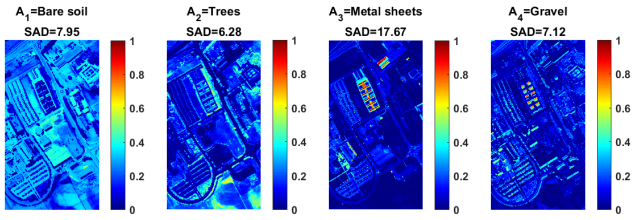


(b) Components of $\boldsymbol{B}$

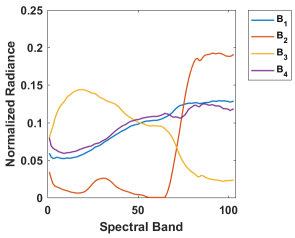


(c) Components of $\boldsymbol{\Psi}$

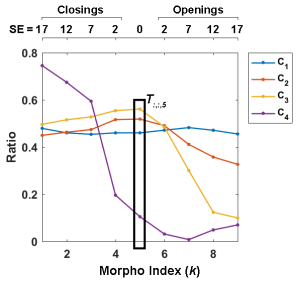
Fig. 11: Pavia. CPD results of the Patch-tensor for R=4



(a) Components of $\boldsymbol{A}$

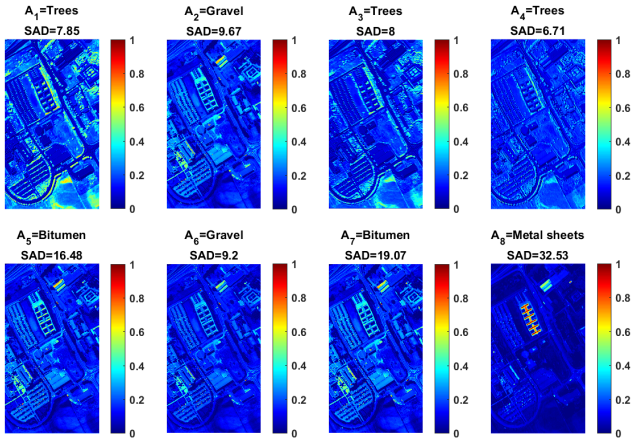


(b) Components of $\boldsymbol{B}$

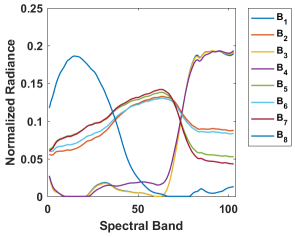


(c) Components of $\boldsymbol{\Psi}$

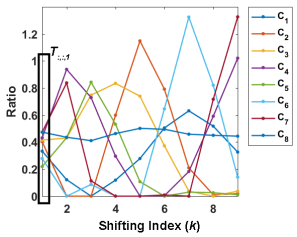
Fig. 12: Pavia. CPD results of the MM-tensor for R=4



(a) Components of $\boldsymbol{A}$

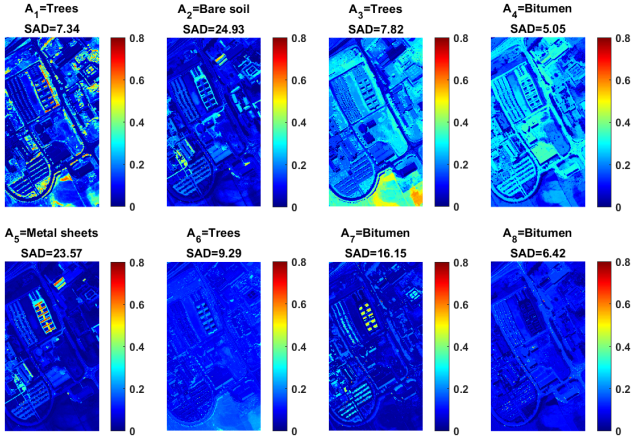


(b) Components of $\boldsymbol{B}$

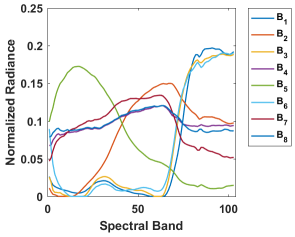


(c) Components of $\boldsymbol{\Psi}$

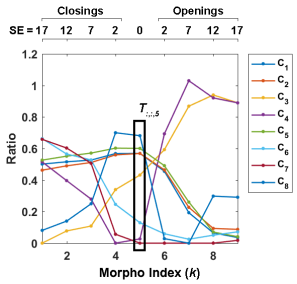
Fig. 13: Pavia. CPD results of the Patch-tensor for R=8



(a) Components of $\boldsymbol{A}$



(b) Components of $\boldsymbol{B}$



(c) Components of $\boldsymbol{\Psi}$

Fig. 14: Pavia. CPD results of the MM-tensor for R=8

2) Few latent components, ELMM and SV: Here, we are interested in the property of CPD dealing with SV. Since Patch-HSI tensors have as an inherently low structure as that of NMF, we start by considering $R = 4$ for patches and MM. As explained in Section IV-B, such a low value highlights the SV and *spatial regularization* aspects of MultiHU-TD. We compile the discussion into two stages: (a) Fig. 11 representing Patch-CPD, and (b) Fig. 12 representing MM-CPD.

2a) In Fig. 11, $\boldsymbol{B}_1$ and $\boldsymbol{B}_4$ are identified as *Trees* and *Metal Sheets* respectively, which reflects the areas highlighted in $\boldsymbol{A}_1$ and $\boldsymbol{A}_4$, while $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ can be seen as a bundle identified as *Gravel*, which reflects the areas highlighted in $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ and where we already start to see *replicated* components. In fact, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ are *slightly-shifted* versions of each other, which is explained better in the following interpretation of $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Psi}_3$.

Looking at Fig. 11c, which is the most interesting, one might intuitively expect to see horizontal curves since, quan-

titatively, the collective SV is supposedly *constant* in patches. However, while $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_4$ look almost straight, $\mathbf{\Psi}_2$ and $\mathbf{\Psi}_3$ are not. We notice here that where $\mathbf{\Psi}_2$ is high, $\mathbf{\Psi}_3$ is low and vice versa. In part, this means that $\boldsymbol{A}_2$ represents the shifts where $k = \{2, 3, 4, 5\}$, while $\boldsymbol{A}_3$ represents those where $k = \{6, 7, 8, 9\}$. In another part, the two columns fluctuate in a way that *maintains* a constant SV and *balances out* their quantitative presence across the frontal slices. As for $k = 1$, which represents the original HSI, we notice that the scaling factors are almost equal, which means that the spectral vectors of the convex hull are *equally present* in the HSI, all of which shows that Patch-HSI tensors do not account to physical spatial effects.

2b) In Fig. 12, $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, and $\boldsymbol{B}_3$ are identified as *Bare Soil*, *Trees*, and *Metal Sheets* respectively, which reflects the areas highlighted in $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and $\boldsymbol{A}_3$, all of which is similar to those obtained by patches. As for $\{\boldsymbol{A}_4, \boldsymbol{B}_4\}$, while $\boldsymbol{B}_4$ and $\boldsymbol{B}_1$ can be seen as a spectral bundle, unlike patches, we notice that $\boldsymbol{A}_4$ highlights interesting *shadow* areas (i.e., dark features), which clearly reflects the *morphological awareness* incorporated into CPD with MM. The latter becomes more interesting with the following interpretation of $\mathbf{\Psi}$.

Looking at Fig. 12c, we observe three main patterns that can be associated to the chosen morphological parameters. First, $\mathbf{\Psi}_4$ corresponds to dark features (reflected by the shadows in $\boldsymbol{A}_4$) as it has higher values when k corresponds to CbR, then continues decreasing towards ObR. Second, $\mathbf{\Psi}_2$ and $\mathbf{\Psi}_3$ correspond to small features as they have higher values around the middle ($k = 5$) where the SEs are small, which is visually reflected through the small objects highlighted in $\boldsymbol{A}_2$ (trees) and $\boldsymbol{A}_3$ (metal sheets and vehicles). Third, $\mathbf{\Psi}_1$ is rather steady, which means that the spatial features shown in $\boldsymbol{A}_1$ are general.

As for $k = 5$, which represents the original HSI, we notice that $\mathbf{\Psi}_2$ and $\mathbf{\Psi}_3$ have the highest scaling factors since they correspond to relatively bright objects of the scene, $\mathbf{\Psi}_1$ has a slightly lower factor since it corresponds to darker objects like asphalt roads, building roofs, parking lots, and bare soil areas, and $\mathbf{\Psi}_4$ has the lowest factors since it corresponds to dark shadows. These relationships show the column- and row-wise significance of $\mathbf{\Psi}$ and how MultiHU-TD can balance out the SVs and simultaneously reconstruct the original HSI.

3) More latent components, ELMM and multi-feature separability: Here, we are interested in MultiHU-TD when we have more degrees of freedom, where we dive deeper into the factors of patches and MM for $R = 8$. As explained in Section IV-B, such a higher value demonstrates the multi-feature separability of MM and how patches only replicates its components. We compile the discussion into two stages: (a) Fig. 13 representing Patch-CPD, and (b) Fig. 14 representing MM-CPD.

3a) In Fig. 13, we end up with more replicas of the same $\boldsymbol{A}$ and $\boldsymbol{B}$ components obtained in Fig. 11, the bundles being associated to the column indices $\{1, 3, 4\}$ detected as *Trees*, and $\{2, 5, 6, 7\}$ detected as *Gravel* and *Bitumen*. As for the plot of $\mathbf{\Psi}$, the same remarks of Fig. 11c about balancing the constant SVs in patches apply on Fig. 13c, but since there are many replicas, the figure becomes hard to read. Finally, we notice again that for $k = 1$, the scaling factors are almost equal.

3b) In Fig. 14, we notice that MultiHU-TD is done based on spectral and morphological properties, where we observe three bundles: $\{1, 3, 6\}$, $\{2, 4, 7, 8\}$, and $\{5\}$. $\boldsymbol{B}_1$, $\boldsymbol{B}_3$, and $\boldsymbol{B}_6$ are identified as vegetation, which respectively reflects the areas highlighted in $\boldsymbol{A}_1$ (small vegetation areas like trees), $\boldsymbol{A}_3$ (big vegetation areas like meadow), and $\boldsymbol{A}_6$ (dark shadows on vegetation areas). Therefore, while $\boldsymbol{B}_1$, $\boldsymbol{B}_3$, and $\boldsymbol{B}_6$ form a bundle, unlike patches, the corresponding AMs highlight interesting features accounting to the scale and brightness of vegetation objects, which we discuss in more depth with the interpretation of $\mathbf{\Psi}$ and which applies to the other components as well. $\boldsymbol{B}_2$, $\boldsymbol{B}_4$, $\boldsymbol{B}_7$, and $\boldsymbol{B}_8$ are identified as *Bitumen* and *Gravel*, which respectively reflects the areas highlighted in $\boldsymbol{A}_2$ (small or short areas of roads and roofs), $\boldsymbol{A}_3$ (big connected areas of roads, roofs, and parking lots), $\boldsymbol{A}_7$ (dark shadows on parking lots and buildings), and $\boldsymbol{A}_8$ (tiny bright vehicles). Finally, $\boldsymbol{B}_4$ is identified as *Metal Sheets*, which are small.

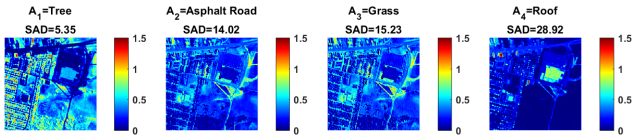
Looking at Fig. 14c, we observe four main patterns that can be associated to the chosen morphological parameters: First, $\mathbf{\Psi}_6$ and $\mathbf{\Psi}_7$ correspond to dark features (as observed in $\boldsymbol{A}_6$ and $\boldsymbol{A}_7$) as they have higher values when k corresponds to CbR, then continue decreasing towards ObR. Second, $\mathbf{\Psi}_1$, $\mathbf{\Psi}_2$, and $\mathbf{\Psi}_5$ correspond to small features (as observed in $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and $\boldsymbol{A}_4$) as they have higher values around $k = 5$ where the SEs are small. Third, $\mathbf{\Psi}_3$ and $\mathbf{\Psi}_4$ correspond to big features (as observed in $\boldsymbol{A}_3$ and $\boldsymbol{A}_4$) as they have higher values when k corresponds to big SEs with ObR. Fourth, $\mathbf{\Psi}_8$ corresponds to the tiny vehicles as it is the highest when k corresponds to the smallest SE. Finally, we talk about the original HSI in the following (i.e., $\boldsymbol{T}_{:,:,5}$).

As for $k = 5$, we notice that $\mathbf{\Psi}_1$, $\mathbf{\Psi}_2$, $\mathbf{\Psi}_5$, and $\mathbf{\Psi}_8$ have the highest scaling factors (relatively bright objects including the vehicles), $\mathbf{\Psi}_3$ and $\mathbf{\Psi}_4$ have lower factors (darker objects like asphalt roads, building roofs, parking lots, bare soil, and meadow areas), and $\mathbf{\Psi}_6$ and $\mathbf{\Psi}_7$ have the lowest factors (dark shadowy features). These relationships showcase the separability of MultiHU-TD when the third-mode has a significant physical meaning and when the number of latent components is set to be sufficiently high, which can also be interpreted in terms of ELMM and balancing the SV factors.

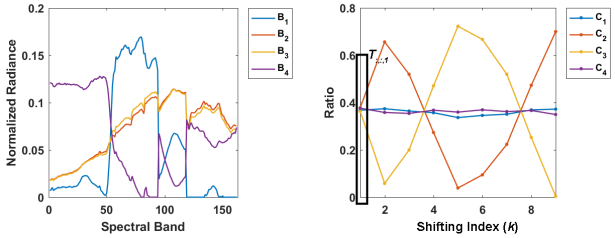
B. Results - Urban

In this section, we present the experiments of the Urban HSI following the same order of Pavia. Since we have the same observations, and in order to avoid repetition, we briefly go over the results. But first, we note that the spatial and spectral references already come with the downloaded dataset, consisting of four endmembers, which we use as a spectral reference (shown in figure 10c) in the experiments, and four abundance maps (shown in figure 10b). For the MM-HSI tensor, our SEs are disks with the successive radii: $\{1, 4, 7, 10\}$ pixels. Both Patch-HSI and MM-HSI tensors then have $K = 9$ frontal slices and dimensions $94249 \times 162 \times 9$. Finally, we also choose $R = 4$ and $R = 8$ for the number of latent components.

1) AO-ADMM-ASC: Table VI shows the RMSE results of MM-CPD between AO-ADMM-ASC and Naive ASC [14],

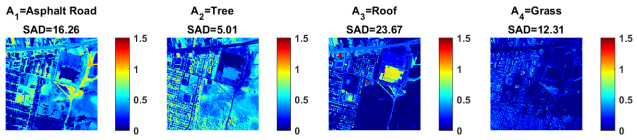


(a) Components of $\boldsymbol{A}$

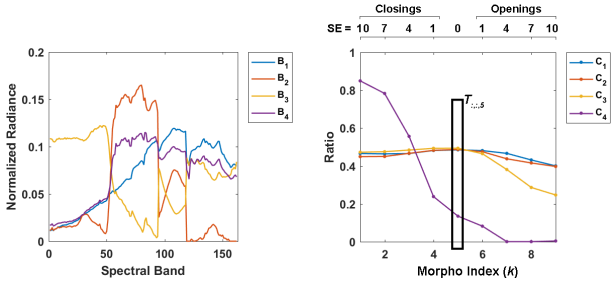


(b) Components of $\boldsymbol{B}$ (c) Components of $\boldsymbol{\Psi}$

Fig. 15: Urban. CPD results of the Patch-tensor for R=4



(a) Components of $\boldsymbol{A}$



(b) Components of $\boldsymbol{B}$ (c) Components of $\boldsymbol{\Psi}$

Fig. 16: Urban. CPD results of the MM-tensor for R=4

where again with AO-ADMM-ASC we gain in RMSE with a small difference in the execution time.

Algorithm	R	RMSE %	Time (s)
Naive ASC [14]	8	7.88	124
AO-ADMM-ASC	8	6.87	251

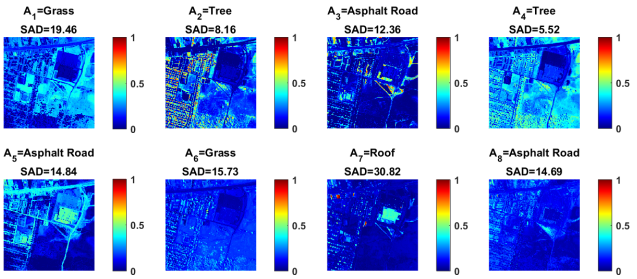
TABLE VI: Urban. The results of AO-ADMM-ASC and Naive ASC in terms of RMSE and execution time: R is the number of latent components. The results of the minimum RMSE are shown.

2) Few latent components, ELMM and SV: Here, we discuss the results for $R = 4$, where (a) Fig. 15 represents Patch-CPD (b) Fig. 16 represents MM-CPD. A fast look at the figures shows that we have the same observations as those of Pavia:

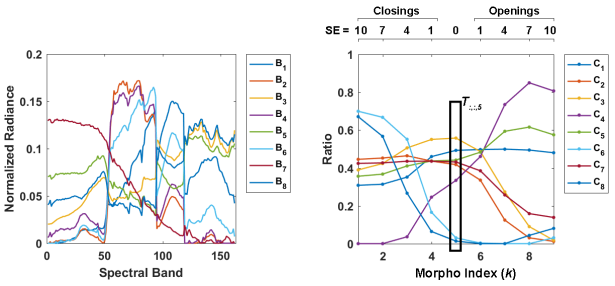
2a) In Fig. 15, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ form a bundle, and $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ are *replicas* and represent *Asphalt+Grass*. Moreover, in Fig. 15c, we see the same patterns and fluctuations that were observed in Fig. 11c related to the constant SV and its quantitative balance in patches, and the scaling factors are equal for $k = 1$. On the other hand, $\{\boldsymbol{A}_1, \boldsymbol{B}_1\}$ and $\{\boldsymbol{A}_4, \boldsymbol{B}_4\}$ represent *Tree+Grass* and *Roof* respectively with steady $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_4$ patterns.

2b) In Fig. 16, while $\boldsymbol{B}_4$ and $\boldsymbol{B}_1$ form a bundle, we notice that $\boldsymbol{\Psi}_4$ has the same pattern observed in Fig. 12c, which corresponds to dark shadows and is reflected in $\boldsymbol{A}_4$, which highlights shadows of buildings and trees that fall on grass areas. As for the other components, they can be interpreted similarly to those in the case of Pavia (including for $k = 5$), where $\{\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{\Psi}_1\}$, $\{\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{\Psi}_2\}$, and $\{\boldsymbol{A}_3, \boldsymbol{B}_3, \boldsymbol{\Psi}_3\}$ represent *Asphalt+Grass*, *Tree+Grass*, and *Roof* respectively.

3) More latent components, ELMM and multi-feature separability: Here, we discuss the results for $R = 8$, where Fig. 17 represents MM-CPD. We skip the case of Patch-CPD in order to avoid repetition, where we simply end up with more *replicas* of the components of figures 15a and 15b. In Fig. 17, we are interested in the features of the AMs that do not appear in Patch-CPD as the comments on the spectral and morphological



(a) Components of $\boldsymbol{A}$



(b) Components of $\boldsymbol{B}$ (c) Components of $\boldsymbol{\Psi}$

Fig. 17: Urban. CPD results of the MM-tensor for R=8

patterns are the same as those of Pavia; where the plots reflect the qualitative features that appear in the respective AMs.

We observe three column sets forming three spectral bundles: $\{1, 2, 6\}$, $\{3, 5, 7, 8\}$, and $\{4\}$. First, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_4$, and $\boldsymbol{A}_6$ were identified as *Vegetation*: $\boldsymbol{A}_1$ highlights grass fields, which is close to the *Grass* reference and does not appear in Patch-CPD. $\boldsymbol{A}_2$ and $\boldsymbol{A}_4$ highlight small and big areas, and together they correspond to the *Tree* reference. $\boldsymbol{A}_6$ highlights dark shadows (which is reflected in $\boldsymbol{\Psi}_6$). Second, $\boldsymbol{A}_3$, $\boldsymbol{A}_5$, and $\boldsymbol{A}_8$ are identified as *Asphalt Road*: $\boldsymbol{A}_8$ seems to correspond to dark features (refer to $\boldsymbol{\Psi}_8$). $\boldsymbol{A}_3$ highlights small roads such as dirt and narrow streets, while $\boldsymbol{A}_5$ highlights large roads like the main and connected roads, which are clearly highlighted unlike the case of patches. Third, $\boldsymbol{A}_7$ is identified as *Roof* and highlights both small and large building roofs.

VI. Conclusion

In this paper, we proposed a methodological framework for MultiHU-TD based on CPD and the AO-ADMM-ASC algorithm, where the samples (pixels) represent a convex combination of the sources. We also established a unified framework for the interpretability of MultiHU-TD into "multilinear" subspaces which involved mathematical, physical, and graphical representations of the CPD model with ASC, ELMM and SV. Finally, we proposed to include MM as spatial features in a spectral-spatial HU and dived further into the case of neighborhood patches, where MM incorporates physically meaningful features into the data tensor. Through the comparison between the two third-mode examples, we provided in-depth insights on the interpretability of MultiHU-TD including the physical significance of the factor matrices and the input rank. To conclude, we summarize some key properties of MultiHU-TD as follows:

- Multi-feature hyperspectral data is useful for low-rank latent variable analysis, such as unmixing.
- Having multiple modalities of features allows to exploit more information on the scene, relaxing the dependency on the high-rank spatial structures while conserving enough context of the scene.
- Having multiple modalities of features with CPD acts as an implicit prior on the scene. The MultiHU-TD framework is then equivalent to performing a coupled matrix decomposition on each of the tensor slices where the abundances matrix $\boldsymbol{A}$ is the common factor.
- Multi-feature unmixing with low-rank tensor decomposition factorizes the pixel and spectral information and implicitly models the spectral variabilities of the scene.

In the future, we plan to explore BTD which allows some flexibility with the tensor structure and can be seen as an extension to Spectral Bundles for SV [11], but also comes with many challenges such as the rank and the interpretation of the subspaces. Moreover, areas of BSS other than HU may be explored. Finally, it is worth mentioning that some deep learning approaches are being considered for HU (which still suffer from the increasing and flexible dimensionality of HSIs and the difficulty of finding data sets for training especially in a blind framework). However, by developing our methodological study of tensor-based unmixing and pushing for interpretability, we pave the way towards tensor-based interpretable and hybrid deep learning models as well as tensor subspace learning [21], [59], [60], which can be very helpful in blind settings such as BSS and HU.

VII. Acknowledgments

We would like to thank the associate editor and the reviewers for their valuable comments and for enriching the content and the context of this work.

Appendix A
Acronyms

Matrix & Tensor Decomposition

ADMM	alternating direction method of multipliers
ALS	alternating least squares
AO	alternating optimization
AO-ADMM	alternating optimization ADMM
ASC	abundance sum-to-one constraint
BSS	blind source separation
BTD	block term decomposition
CPD	canonical polyadic decomposition
NMF	nonnegative matrix factorization
ProCo-ALS	projected compressed ALS

Hyperspectral Imaging

AM	abundance map
ELMM	extended linear mixing model
EM	endmember
GT	ground-truth
HSI	hyperspectral image
HU	hyperspectral unmixing
LiDAR	light detection and ranging
LMM	linear mixing model
MultiHU-TD	multi-feature hyperspectral unmixing based on tensor decomposition
RMSE	root mean squared error
SAD	spectral angular distance
SNR	signal-to-noise ratio
SV	spectral variability

Mathematical Morphology

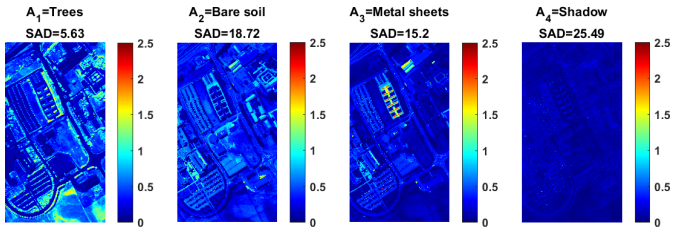
CbR	closing by reconstruction
EMP	extended morphological profile
MM	mathematical morphology
MP	morphological profile
ObR	opening by reconstruction
SE	structuring element

Appendix B
Sparse-NMF results on the matricized HSI

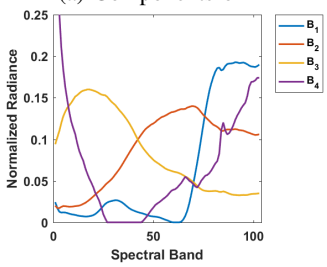
In this appendix, we include the results obtained by applying sparse NMF (with ASC) [17], which partly inspired this work. We note that these results cannot be compared with those of tensor decomposition in terms of RMSE and execution time due to the following reasons:

- In terms of RMSE, on the one hand, we reconstruct a multi-feature HSI tensor, while on the other hand, we reconstruct a HSI matrix which does not apply in *multi-feature analysis*, so the reconstructed data represent different types of information.
- In terms of execution time, NMF typically has shorter execution times than tensor-based methods due to the added complexity. However, both tools are fundamentally different and can not be used for the same multi-linear application.

With that said, the obtained results serve only as a qualitative baseline or reference for the *abundance maps* and *spectral components* of decomposing the two HSI datasets, which can indeed be compared with those obtained in the case of MM-HSI and Patch-HSI tensors.

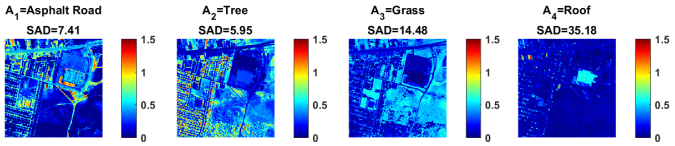


(a) Components of $\boldsymbol{A}$

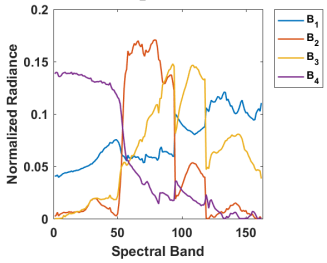


(b) Components of $\boldsymbol{B}$

Fig. 18: Pavia. NMF results of the HSI matrix for $R = 4$



(a) Components of $\boldsymbol{A}$



(b) Components of $\boldsymbol{B}$

Fig. 19: Urban. NMF results of the HSI matrix for $R = 4$

A. Pavia dataset

Fig. 18 shows the results obtained for the dataset of Pavia University. Each abundance map is shown with the class that is assigned to it based on the minimum SAD value, which is reported as well.

First, we look at the first three components, *Trees*, *Bare Soil*, and *Metal Sheets*. Their spectral signatures $\{\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3\}$ shown in Fig. 18b look very similar to those of the reference, but the corresponding SAD values are relatively bad (high) compared to those obtained using tensor decomposition in Fig. 11 and 12.

The abundance maps $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3\}$ of Fig. 18a show highlighted elements belonging to *Trees*, *Bare Soil*, and *Metal Sheets* respectively. However, we notice that other areas of the scene belonging to these categories are barely or faintly highlighted (e.g., asphalt road, brick parking lots, other soil areas), which is due to the insufficiency of LMM to model their variabilities.

Regarding the *Shadow* component, the spectral signature $\boldsymbol{B}_4$ looks slightly similar to the reference, but the shadows in the map are barely visible due to their relatively very low brightness $\boldsymbol{A}_4$.

B. Urban dataset

Fig. 19 shows the abundance maps and spectral signatures of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. We obtain four components with relatively low (good) SAD values and good abundance maps similarity with respect to the reference. Finally, we also note that some dark areas like asphalt roads are not highlighted.

APPENDIX C
SYNTHETIC HSI EXAMPLE

In the following, we simulate a time-series HSI tensor through a synthetic example. The noiseless tensor is reconstructed from its building blocks, i.e., the factor matrices, which are considered here as the noiseless ground-truths. This simplified example serves as a demonstration of the performance of MultiHU-TD as we increase the signal-to-noise ratio (SNR).

1) Description of the synthetic dataset: The synthetic HSI cube $\boldsymbol{\mathcal{M}}$ of a signle time-stamp has dimensions $128 \times 128 \times 26$, whose matricized version is denoted by $\boldsymbol{M}$ of dimensions 16384×26. In particular:

- The spatial scene is composed of six objects that vary in size, as shown in Fig. 20, denoted by "Object 1" through "Object 6". The scene is shown in Fig. 22a in false colors.
- Three independent spectral signatures are simulated from the HSI of Pavia University, corresponding to `Street`, `Vegetation`, and `Metal Sheets` which are shown in Fig. 21b.
- The objects are assigned linear mixtures of the three simulated spectra as shown in TABLE VII. This means that the ground-truth of the abundances satisfies the ASC.

That said, $\boldsymbol{M}$ has rank 3 where $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{B}^{\mathsf{T}}$, such that $\boldsymbol{A} \in \mathbb{R}^{16384\times 3}$ and $\boldsymbol{B} \in \mathbb{R}^{26\times 3}$ are two factor matrices. The spatial and spectral ground-truths, of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively, are shown in Fig. 21a and 21b.

Object number	1	2	3	4	5	6
`Street` (%)	10	0	0	**80**	20	**100**
`Vegetation` (%)	**70**	**100**	0	10	20	0
`Metal Sheets` (%)	20	0	**100**	10	**60**	0
`Total` (%)	100	100	100	100	100	100

TABLE VII: Spectral mixture of each object based on the end-members, `Street`, `Vegetation`, and `Metal Sheets`.

As for the full time-series HSI tensor, it is composed of three stamps where the objects of the scene change in color or disappear in time. In principle, this corresponds to a time-series tensor $\boldsymbol{\mathcal{D}}$ of dimensions $128 \times 128 \times 26 \times 3$. After reordering the pixel modalities in lexicographic order, we would obtain a tensor $\boldsymbol{\mathcal{T}}$ of dimensions $16384 \times 26 \times 3$. We synthesize the time-series HSI tensor $\boldsymbol{\mathcal{T}}$ of dimensions $\mathbb{R}^{16384\times 26\times 3}$ from the product with an additional matrix $\boldsymbol{C} \in \mathbb{R}^{3\times 3}$ such that:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{I}} \underset{1}{\bullet} \boldsymbol{A} \underset{2}{\bullet} \boldsymbol{B} \underset{3}{\bullet} \boldsymbol{C} \tag{32}$$

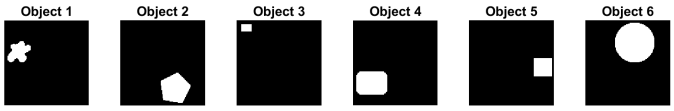
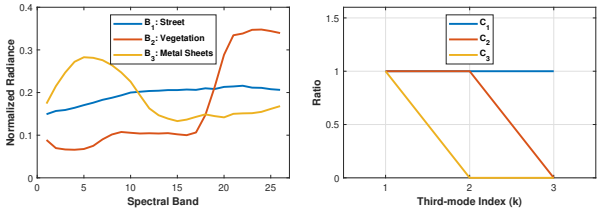


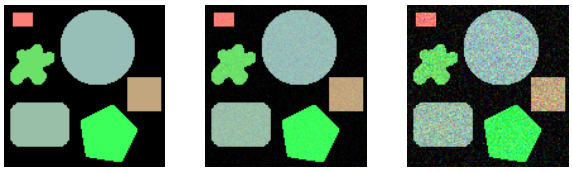
Fig. 20: The simulated spatial objects of the synthetic HSI, denoted as Objects "1" through "6".

(a) Ground-truth of $\boldsymbol{A}$

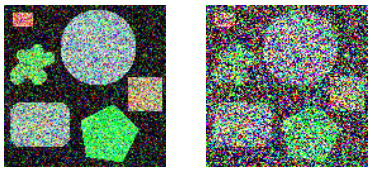


(b) Ground-truth of $\boldsymbol{B}$ (c) Ground-truth of $\boldsymbol{C}$

Fig. 21: Ground-truth components of the synthetic tensor $\boldsymbol{\mathcal{T}}$.



(a) Noiseless (b) $\sigma^2 = 10^{-4}$ (c) $\sigma^2 = 10^{-3}$



(d) $\sigma^2 = 10^{-2}$ (e) $\sigma^2 = 10^{-1}$

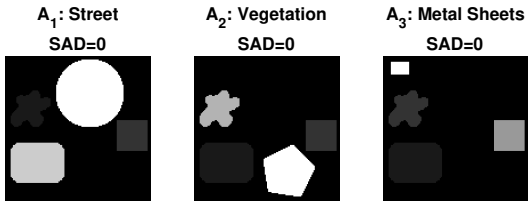
Fig. 22: The synthetic HSI scene at a single time stamp $k = 1$ in False colors, with Gaussian noise of variance σ^2 ranging from 0 (being noiseless) to 10^{-1} (being very noisy).

where $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{3\times3\times3}$ is a diagonal tensor of ones, and $\boldsymbol{C}$ is described as follows:

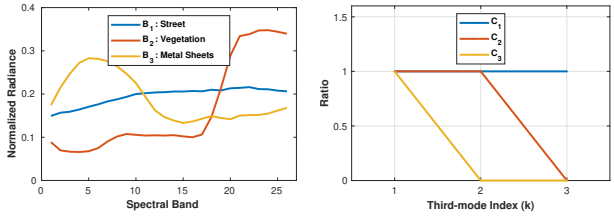
$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} <=> \begin{pmatrix} k=1 \\ k=2 \\ k=3 \end{pmatrix} \tag{33}$$

where k is the index spanning the third modality, which is that of time stamps. C is considered the ground-truth of the temporal signatures, and its columns are plotted in Fig. 21c.

In order to assess the performance of the proposed framework in the presence of noise, Gaussian noise was added on the tensor $\boldsymbol{\mathcal{D}}$ with varying levels of noise, where the variance $\sigma^2 \in \{0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, as described in Fig. 22. After adding noise, the tensor is reshaped back to $\boldsymbol{\mathcal{T}}$.

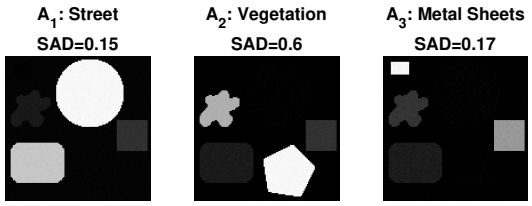


(a) Recovered columns of $\boldsymbol{A}$

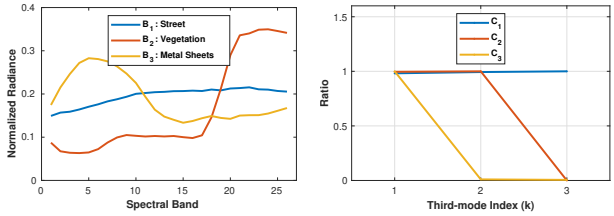


(b) Recovered columns of $\boldsymbol{B}$ (c) Recovered columns of $\boldsymbol{C}$

Fig. 23: Synthetic HSI. Results of decomposing the HSI tensor $\boldsymbol{\mathcal{T}}$ without noise, where we also have RMSE $= 0$.



(a) Recovered columns of $\boldsymbol{A}$

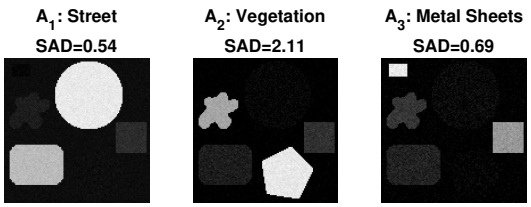


(b) Recovered columns of $\boldsymbol{B}$ (c) Recovered columns of $\boldsymbol{C}$

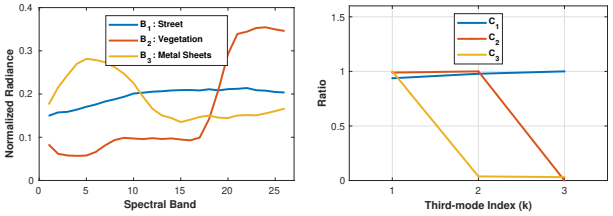
Fig. 24: Synthetic HSI. Results of decomposing the HSI tensor $\boldsymbol{\mathcal{T}}$ with Gaussian noise of variance 10^{-4}.

2) Results of unmixing without adding noise: Fig. 23 shows the components of the factor matrices after decomposing the synthetic time-series HSI tensor described using CP decomposition. The components of the factor matrices are perfectly recovered thanks to the CP uniqueness of the data tensor [61]. Moreover, it is worth noting that the RMSE between the original data tensor and the reconstructed one is 0, which means perfect reconstruction.

3) Results of unmixing with varying levels of noise: In this case, we discard the reconstructability of the tensor itself and focus on the factor matrices, of which we have the abundance matrix $\boldsymbol{A}$. Fig. 24, Fig. 25, Fig. 26, and Fig. 27 show the results of decomposing the tensor under varying levels of Gaussian noise, with variances of 10^{-4}, 10^{-3}, 10^{-2}, and 10^{-1} respectively. Moreover, Fig. 28 shows the evolution of the SAD index of the estimated endmembers as the noise level increases. We note that the experiments are done without applying any spatial denoising.

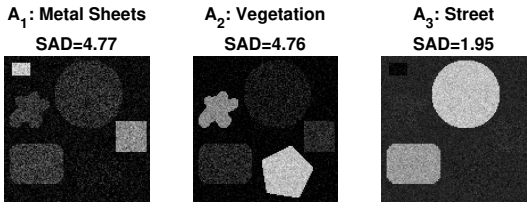


(a) Recovered columns of $\boldsymbol{A}$

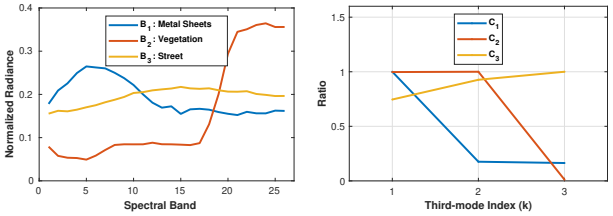


(b) Recovered columns of $\boldsymbol{B}$ (c) Recovered columns of $\boldsymbol{C}$

Fig. 25: Synthetic HSI. Results of decomposing the HSI tensor $\boldsymbol{\mathcal{T}}$ with Gaussian noise of variance 10^{-3}.



(a) Recovered columns of $\boldsymbol{A}$

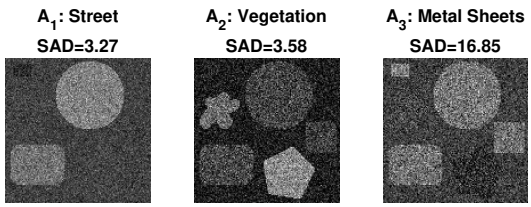


(b) Recovered columns of $\boldsymbol{B}$ (c) Recovered columns of $\boldsymbol{C}$

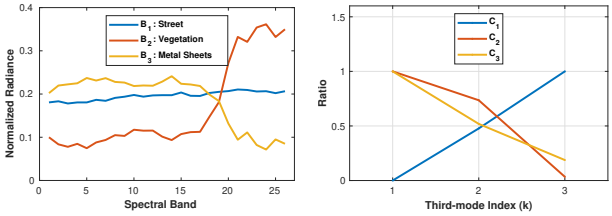
Fig. 26: Synthetic HSI. Results of decomposing the HSI tensor $\boldsymbol{\mathcal{T}}$ with Gaussian noise of variance 10^{-2}.

The components of the factor matrices, including the abundance matrix $\boldsymbol{A}$ whose rows sum to one, are quite recoverable up to a level of noise of variance $\sigma^2 = 10^{-3}$. In the case of $\sigma^2 = 10^{-2}$, the components are still recoverable even though the tensor looks quite noisy in Fig. 22d. In the case of $\sigma^2 = 10^{-1}$, which is very noisy that some of the objects are indistinguishable in Fig. 22e, the reconstructed spectra are still fairly close to their ground-truths and the spatial structures in the factors of the abundance matrix can still be recognized. In Fig. 28, we can see that the SAD index generally increases with the level of noise. However, the values remain quite small, i.e., less than $5°$, indicating the recoverability of the estimated endmembers even under such high levels of noise. The only exception in this case is that of *Metal Sheets* under a variance of 10^{-1} which results in a SAD index of about $16°$; this can be due to the small size of *Metal Sheets* objects, which makes them more susceptible to noise.

Moreover, we note that with the addition of noise, the tensor



(a) Recovered columns of $\boldsymbol{A}$



(b) Recovered columns of $\boldsymbol{B}$ (c) Recovered columns of $\boldsymbol{C}$

Fig. 27: Synthetic HSI. Results of decomposing the HSI tensor $\boldsymbol{\mathcal{T}}$ with Gaussian noise of variance 10^{-1}.

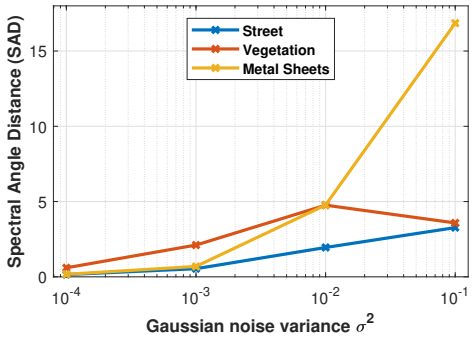


Fig. 28: Synthetic HSI. Evolution of SAD index of the estimated spectra compared to the spectral ground-truth.

$\boldsymbol{\mathcal{T}}$ which was synthesized to be of rank $R = 3$ becomes full rank. Decomposing the noisy tensor with a low rank can be roughly seen as a denoising procedure since it forces the projection of the data onto a lower-rank multi-linear latent subspace of rank $R = 3$. However, spatial denoising is still needed in order to recover a better representation of the spatial components, but said application is out of the scope of this paper.

Finally, this is a simple, controlled, and minimalistic example that serves as an intuition for more complex structures where the situation is completely blind, such as in real HSI tensors.

APPENDIX D
COMPUTATIONAL COMPLEXITY OF AO-ADMM-ASC

First, we refer to paper [49] for the detailed explanations concerning the CPD by using AO-ADMM. Let us consider a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ (where $I_1 >>> I_2 I_3$) with a low rank R, and we consider the complexity as per ADMM iterations. Also, it is important to note that matrices $\tilde{\boldsymbol{W}}$ and $\boldsymbol{T}$ are independent from the inner-ADMM updates, so they can be used only once to compute the products $\tilde{\boldsymbol{W}}^{\mathsf{T}}\tilde{\boldsymbol{W}}$

and $\tilde{\boldsymbol{W}}^{\mathsf{T}}\boldsymbol{T}$, whose values can be cached before the ADMM update allowing to save a lot of repetitive computations. Now, we split the problem into three steps:

1) Unconstrained CPD: In this case, the complexity of the algorithm is dominated only by the updates of the factor matrices. Hence, the complexity is $\boldsymbol{\mathcal{O}}(I_d R^2)$ $\forall d \in \{1,2,3\}$.
2) Nonnegative CPD, which is relevant for the updates of each of the factor matrices: Nonnegativity requires only element-wise projection, i.e., a complexity of $\boldsymbol{\mathcal{O}}(I_d R)$ $\forall d \in \{1,2,3\}$, which is negligible compared to $\boldsymbol{\mathcal{O}}(I_d R^2)$. Hence, the complexity is still dominated by $\boldsymbol{\mathcal{O}}(I_d R^2)$.
3) Nonnegative CPD with sparsity and ASC, which is only relevant for the update of $\boldsymbol{A}$: *Sparsity* is like nonnegativity as it boils down to an element-wise subtraction with complexity $\boldsymbol{\mathcal{O}}(I_1 R)$, which is negligible compared to $\boldsymbol{\mathcal{O}}(I_1 R^2)$. As for *ASC*, it requires two updates:
   - $b_{J+1,r} = \delta \psi_{K,r}^{-1}$ $\forall r \in \{1,\dots,R\}$, i.e., a complexity of $\boldsymbol{\mathcal{O}}(R)$, which is negligible.
   - $t_{i,J+1,k} = \sum_{r=1}^{R} a_{i,r} b_{J+1,r} \psi_{k,r}$ $\forall i \in \{1,\dots,I\}$ and $\forall k \in \{1,\dots,K-1\}$, i.e., a complexity of $\boldsymbol{\mathcal{O}}(I_1 I_3 R)$.

Considering that MultiHU-TD admits a *low-rank* CP decomposition, R is usually small, and in most of the cases we would have $R < I_3$ (or at least very close). In which case, the complexity is dominated by $\boldsymbol{\mathcal{O}}(I_1 I_3 R)$, which changes linearly with either the number of pixels I_1, the third-mode features I_3, or the latent components R.

REFERENCES

[1] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, 2013.

[2] J. Amigo, *Hyperspectral Imaging*, ser. ISSN. Elsevier Science, 2019.

[3] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Oxford UK, Burlington USA: Academic Press, 2010, hal-00460653.

[4] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 34–43, 2014.

[5] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[6] T. Bajjouk, I. Zarati, L. Drumetz, and M. Dalla Mura, "Spatial characterization of marine vegetation using semisupervised hyperspectral unmixing," in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2019, pp. 1–5.

[7] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "A low-rank tensor regularization strategy for hyperspectral unmixing," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 373–377.

[8] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.

[9] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Improved hyperspectral unmixing with endmember variability parametrized using an interpolated scaling tensor," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2177–2181.

[10] S. G. Azar, S. Meshgini, S. Beheshti, and T. Y. Rezaii, "Linear mixing model with scaled bundle dictionary for hyperspectral unmixing with spectral variability," *Signal Processing*, p. 108214, 2021.

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

[12] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3890–3905, 2016.

[13] L. Drumetz, M. Dalla Mura, G. Tochon, and R. Fablet, "Learning endmember dynamics in multitemporal hyperspectral data using a state-space model formulation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2483–2487.

[14] M. A. Veganzones, J. E. Cohen, R. C. Farias, K. Usevich, L. Drumetz, J. Chanussot, and P. Comon, "Canonical polyadic decomposition of hyperspectral patch tensors," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 2176–2180.

[15] M. Jouni, "Image analysis based on tensor representations," Ph.D. dissertation, Université Grenoble Alpes, 2021.

[16] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.

[17] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1112–1125, 2010.

[18] F. Xiong, Y. Qian, J. Zhou, and Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[19] Y. Tan, Z. Li, Y. Xiao, and N. Liu, "Infrared small target detection algorithm based on robust tensor decomposition model within bayesian framework," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1160–1163.

[20] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6868–6877.

[21] B. B. Gatto, E. M. dos Santos, A. L. Koerich, K. Fukui, and W. S. Junior, "Tensor analysis with n-mode generalized difference subspace," *Expert Systems with Applications*, vol. 171, p. 114559, 2021.

[22] M. Zare, M. S. Helfroush, K. Kazemi, and P. Scheunders, "Hyperspectral and multispectral image fusion using coupled non-negative tucker tensor decomposition," *Remote Sensing*, vol. 13, no. 15, p. 2930, 2021.

[23] J. Yao, D. Hong, L. Xu, D. Meng, J. Chanussot, and Z. Xu, "Sparsity-enhanced convolutional decomposition: A novel tensor-based paradigm for blind hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[24] M. A. Veganzones, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, "Nonnegative tensor cp decomposition of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2577–2588, 2016.

[25] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.

[26] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled tucker approximation: Recoverability and svd-based algorithms," *IEEE Transactions on Signal Processing*, 2020.

[27] K. Uto, M. D. Mura, and J. Chanussot, "Normal direction and true color estimation of leaves based on tensor decomposition of leaf-scale optical images," in *Whispers 2018 - Poster Session*, 2018.

[28] Z. Xue, S. Yang, H. Zhang, and P. Du, "Coupled higher-order tensor factorization for hyperspectral and lidar data fusion and classification," *Remote Sensing*, vol. 11, no. 17, p. 1959, 2019.

[29] M. Jouni, M. Dalla Mura, and P. Comon, "Classification of hyperspectral images as tensors using nonnegative CP decomposition," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2019, pp. 189–201.

[30] ——, "Hyperspectral image classification using tensor CP decomposition," in *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, Yokohama, Japan, July 28 - August 2, 2019*, 2019, pp. 1164–1167.

[31] ——, "Hyperspectral image classification based on mathematical morphology and tensor decomposition," *Mathematical Morphology - Theory and Applications*, vol. 1, 2019.

[32] Y. Gu, T. Liu, and J. Li, "Superpixel tensor model for spatial-spectral classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[33] Y. Mitsufuji, N. Takamune, S. Koyama, and H. Saruwatari, "Multichannel blind source separation based on evanescent-region-aware nonnegative tensor factorization in spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 607–617, 2020.

[34] M. Niknazar, H. Becker, B. Rivet, C. Jutten, and P. Comon, "Blind source separation of underdetermined mixtures of event-related sources," *Signal Processing*, vol. 101, pp. 52–64, 2014.

[35] J. Virta and K. Nordhausen, "Blind source separation of tensor-valued time series," *Signal Processing*, vol. 141, pp. 204–216, 2017.

[36] H. Becker, L. Albera, P. Comon, R. Gribonval, F. Wendling, and I. Merlet, "Brain-source imaging: From sparse to tensor models," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 100–112, 2015.

[37] H. Becker, P. Comon, L. Albera, M. Haardt, and I. Merlet, "Multiway space–time–wave-vector analysis for eeg source separation," *Signal Processing*, vol. 92, no. 4, pp. 1021–1031, 2012.

[38] J. Sole-Casals, C. F. Caiafa, Q. Zhao, and A. Cichocki, "Brain-computer interface with corrupted eeg data: a tensor completion approach," *Cognitive Computation*, vol. 10, no. 6, pp. 1062–1074, 2018.

[39] Z. Zhang, G. I. Allen, H. Zhu, and D. Dunson, "Tensor network factorizations: Relationships between brain structural connectomes and traits," *Neuroimage*, vol. 197, pp. 330–343, 2019.

[40] P. Mishra, J. M. Roger, D. Jouan-Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, and D. N. Rutledge, "Recent trends in multi-block data analysis in chemometrics for multi-source data integration," *TrAC Trends in Analytical Chemistry*, p. 116206, 2021.

[41] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, pp. 44–53, May 2014, hal-00923279.

[42] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l\_\{1/2\}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, 2011.

[43] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, 2014.

[44] X. Xu, J. Li, S. Li, and A. Plaza, "Generalized morphological component analysis for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2817–2832, 2019.

[45] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2011.

[46] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3219–3232, 2016.

[47] J. Cohen, R. Farias, and P. Comon, "Fast decomposition of large nonnegative tensors," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 862–866, 2015.

[48] M. Jouni, M. Dalla Mura, and P. Comon, "Some issues in computing the cp decomposition of nonnegative tensors," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2018, pp. 57–66.

[49] K. Huang, N. Sidiropoulos, and A. Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5052–5065, 2016.

[50] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[51] Y. Qi, P. Comon, and L. H. Lim, "Uniqueness of nonnegative tensor approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2170–2183, Apr. 2016, arXiv:1410.8129.

[52] L. Najman and H. Talbot, *Mathematical morphology: from theory to applications*. John Wiley & Sons, 2013.

[53] P. R. Marpu, M. Pedergnana, M. Dalla Mura, S. Peeters, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *International Journal of Image and Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012.

[54] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, 2010.

[55] M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[56] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester: Wiley, 2009.

[57] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2335–2353, 2014.

[58] M. Dalla Mura, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "The evolution of the morphological profile: From panchromatic to hyperspectral images," in *Optical Remote Sensing*. Springer, 2011, pp. 123–146.

[59] B. B. Gatto *et al.*, "Pattern-set representations using linear, shallow and tensor subspaces," Ph.D. dissertation, Universidade Federal do Amazonas, 2020.

[60] B. Batalo, L. S. Souza, B. B. Gatto, N. Sogi, and K. Fukui, "Temporal-stochastic tensor features for action recognition," *Machine Learning with Applications*, p. 100407, 2022.

[61] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.

Mohamad Jouni (S'14, GS'17, M'23) received the B.Eng. degree in Computer and Communications Engineering from the Lebanese University, Beirut, Lebanon, in 2016, and the M.Sc. and Ph.D. degrees in Signal and Image Processing from Grenoble Institute of Technology and the University of Grenoble Alpes, Grenoble, France, in 2017 and 2021 respectively. In 2019, he was a visiting researcher for 10 weeks at Tokyo Institute of Technology, Tokyo, Japan. Since 2021, he has been a Postdoctoral Researcher at Grenoble Institute of Technology, Grenoble, France. His research interests include computational imaging, tensor algebra, hybrid AI methods, and applications of multimodal and hyperspectral data analysis.

Mauro Dalla Mura (S'08, M'11, SM'18) received the B.Sc. and M.Sc. degrees in Telecommunication Engineering from the University of Trento, Italy in 2005 and 2007, respectively. He obtained in 2011 a joint Ph.D. degree in Information and Communication Technologies (Telecommunications Area) from the University of Trento, Italy and in Electrical and Computer Engineering from the University of Iceland, Iceland. In 2011 he was a Research fellow at Fondazione Bruno Kessler, Trento, Italy, conducting research on computer vision. He is currently an Assistant Professor at Grenoble Institute of Technology (Grenoble INP), France since 2012. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). He is a Junior member of the Institut Universitaire de France (2021-2026). Dr. Dalla Mura has been appointed "Specially Appointed Associate Professor" at the School of Computing, Tokyo Institute of Technology, Japan for 2019-2022. His main research activities are in the fields of remote sensing, computational imaging, image and signal processing. Dr. Dalla Mura was the recipient of the IEEE GRSS Second Prize in the Student Paper Competition of the 2011 IEEE IGARSS 2011 and co-recipient of the Best Paper Award of the International Journal of Image and Data Fusion for the year 2012-2013 and the Symposium Paper Award for IEEE IGARSS 2014. Dr. Dalla Mura was the IEEE GRSS Chapter's Committee Chair for 2020-2021. He was President of the IEEE GRSS French Chapter 2016-2020 (he previously served as Secretary 2013-2016). In 2017 the IEEE GRSS French Chapter was the recipient of the IEEE GRSS Chapter Award and the "Chapter of the year 2017" from the IEEE French Section. He is on the Editorial Board of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS) since 2016.

Lucas Drumetz (Member, IEEE) received the M.Eng. degree form Grenoble INP in 2013, and the Ph.D degree in 2016 in image and signal processing from Université Grenoble Alpes, for works carried out at GIPSA-lab, Grenoble, France. This work has been awarded the PhD award of the University of Grenoble Alpes in 2017. In 2017, he was a Visiting Assistant Professor at the Department of Mathematics at the University of California, Los Angeles (UCLA). In 2017, he was also a visiting researcher for 10 weeks at the RCAST laboratory at the University of Tokyo, Japan. Since He has been an Associate Professor at IMT Atlantique, in the Mathematical and Electrical Engineering department since 2018. He is part of the OSE (Observations, Signal and Environment) team of UMR CNRS 6285 Lab-STICC. His research interests include inverse problems and machine learning for remote sensing applications, signal and image processing, and optimization techniques.

Pierre Comon (M'87 - SM'95 - F'07) received the Graduate degree in 1982, and the Doctorate degree in 1985, both from the University of Grenoble, France. He received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia Antipolis, France, between 1988 and 1997. He was with the ISL laboratory, Stanford University, CA, USA, in 1987. He joined in 1997 the Eurecom Institute, Sophia Antipolis, France. He is a Research Director with CNRS since 1998, first with the Laboratory I3S, Sophia Antipolis, France, until 2012, and then with Gipsa-Lab, Grenoble, France. He has been the Director of Labex Persyval, Grenoble, until 2022. His research interests include high-order statistics (HOS), blind techniques, statistical signal and array processing, tensor decompositions, multi-way factor analysis, and data science. He was an Associate Editor for the IEEE Transactions on Signal Processing from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the Coordinator of the European Basic Research Working Group on HOS, ATHOS, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group. Between 2001 and 2004, he acted as a Launching Associate Editor with the IEEE Transactions on Ciruits and Systems I, in the area of Blind Techniques. He has also been a member of the editorial board of the Elsevier journal Signal Processing from 2006 to 2011, and member of several IEEE Technical Committees. He was in the Editorial Board of the SIAM Journal on Matrix Analysis and Applications from 2011 to 2017. He received several prizes, including the Silver medal of CNRS in 2018. Dr Comon is also a Fellow of Eurasip and SIAM.