NONNEGATIVE BLOCK-TERM DECOMPOSITION WITH THE β-DIVERGENCE: JOINT DATA FUSION AND BLIND SPECTRAL UNMIXING

Clémence Prévost, Valentin Leplat

NONNEGATIVE BLOCK-TERM DECOMPOSITION WITH THE β-DIVERGENCE: JOINT DATA FUSION AND BLIND SPECTRAL UNMIXING

C. Prévost[*]

University of Lille, CNRS
Centrale Lille, UMR 9189 CRIStAL
F-59000 Lille, France

V. Leplat[†]

Skoltech
Center for Artificial Intelligence Technology (CAIT)
Moscow, Russia

ABSTRACT

We present a new method for solving simultaneously two problems:(1) hyperspectral and multispectral image fusion, and (2) the blind spectral unmixing of the unknown super-resolution image. The method, dubbed as β-(L_r,L_r,1)-NBTD, relies on three key elements: (1) the nonnegative decomposition in rank-(L_r,L_r,1) block-terms of the super-resolution tensor, (2) the joint factorization of the input images, and (3) the formulation of a family of optimization problems including the β-divergences objective functions. In order to solve the two problems at hand, we propose multiplicative updates based on majorization-minimization. We come up with a family of simple, robust and efficient algorithms, adaptable to various noise statistics. As a by-product, we propose a new robust initialization for the low-rank block-term factors. We show on numerical experiments that β-(L_r,L_r,1)-NBTD competes favorably with State-Of-The Arts methods for solving the super-resolution problem, while accurately solving the unmixing problem for various noise statistics.

***Index Terms*—** Nonnegative tensor factorization, block-term decomposition, β-divergence, blind spectral unmixing, hyperspectral super-resolution.

I. INTRODUCTION

Hyperspectral devices are able to sample the electromagnetic spectrum into hundred of wavelengths, allowing for the acquisition of hyperspectral images (HSIs) that possess high spectral resolution. However, the tradeoff between spatial and spectral resolution forces the HSIs to have a small number of relatively large pixels [1]. On the other hand, multispectral sensors produce multispectral images (MSIs) with high spatial resolution (smaller pixels), at the cost of a restricted number of spectral bands. The composition of each pixel in HSIs and MSIs can be approximated by a sum of a small number of spectral signatures, or endmembers. This representation is known as the linear mixing model. It allows for so-called blind spectral unmixing, that is, identifying materials present within the scenery with limited prior information, classically by computing the spectral signatures of these materials, usually referred to as endmembers, and their abundance maps.

The hyperspectral super-resolution (HSR) problem [2] was formulated to circumvent the physical limitations of each device. This problem aims at recovering a super-resolution image (SRI) that possesses both high spatial and high spectral resolutions from co-registered HSI and MSI of the same scene. The high spatial and spectral resolutions of the SRI can then be exploited in traditional unmixing tasks. Hence, the goal of developing an efficient method for solving both problems at once can be summarized as follows: identify and localize with higher accuracy the materials present within the scenery at hand by combining multiple information of different resolutions (the MSI and the HSI).

Many approaches have been proposed to solve the HSR problem. Most matrix approaches [3], [4], [5], [6] are based on the linear mixing model and perform a coupled low-rank factorization of the matricized HSI and MSI. Some matrix approaches are suitable for the HSR and unmixing problem as well, which consists of recovering the underlying SRI by means of a physically-informed low-rank approximation. See for instance [7] and [3].

More recently, tensor-based approaches were proposed for the HSR problem. The works of [8], [9] formulate the HSR problem as a coupled canonical polyadic (CP) decomposition, while a coupled multilinear Tucker decomposition is used in [10]. See [11] for an overview of tensor methods for hyperspectral data processing. However, the factors of these decompositions lack physical interpretation, and thus the aforementioned methods cannot be used for unmixing. Motivated by the usefulness of tensor models, approaches based on block-tensor decomposition [12], [13] were proposed for solving the HSR problem. Previous works of the authors [14] used the block-term decomposition for joint fusion and

[*]This work was partly supported by the ANR project "Chaire IA Sherlock" ANR-20-CHIA-0031-01 hold by P. Chainais, as well as by the national support within the *programme d'investissements d'avenir* ANR-16-IDEX-0004 ULNE and Région HDF.

[†]VL acknowledges the support by Ministry of Science and Higher Education grant No. 075-10-2021-068.

unmixing in the presence of spectral variability. A recent work used the block-term decomposition for super-resolution only [15]. This decomposition was also successfully used to perform unmixing [16] on the SRI directly. A multiplicative approach based on majorization-minimization was proposed, but was not suited to coupled tensor models and hence, to super-resolution. These approaches did not fully exploit the flexibility of the block-term decomposition, since they assumed all ranks to be equal, thus resulting in simpler algorithms. Furthermore, they were only suited to isotropic white Gaussian noise.

Contributions: the main contribution of our method is the adequate combination of three key-ingredients, described below.

(1) *Decomposition in rank-(L_r,L_r,1) terms of the SRI*: in this paper, we fully exploit the flexibility of the chosen decomposition by considering ranks possibly different from each other.

(2) *Coupled optimization problems along with nonnegativity constraints*: we propose a family of coupled tensor optimization problems. While we illustrate those problems with an example in remote sensing, other fields of applications can be envisioned for the considered model, such as audio signal processing [17], biomedical imaging [18] or graph signal processing [19]. We handle the nonnegativity constraints using multiplicative updates. To the best of our knowledge, multiplicative algorithms for coupled tensor low-rank models has not been addressed yet at the writing time of this paper.

(3) *β-divergence family as objective functions*: using β-divergences allows us to take into account various noise statistics potentially present within the data, see [20] for a detailed overview of the topic.

Moreover, we develop a family of simple, efficient and flexible algorithms adapted to solve the proposed optimization problems. Compared to previous works, our algorithms have several advantadges: (i) they are adapted to various noise statistics, (ii) they are assorted with convergence guarantees due to the use of multiplicative updates, and (iii) they are able to estimate the degradation operators at stake in the coupled tensor models. While multiplicative matrix-based algorithm was designed in [21] and was able to estimate such matrices, this matter has not been considered yet in the considered application with tensor methods. As a by-product, we provide a robust way to initialize the low-rank factors based on multiplicative updates.

We finally demonstrate that our method competes favorably with the state-of-the-art when applied to solving both problems of interest for synthetic and semi-real data sets including three noise statistics: Gaussian noise, Poisson noise and multiplicative Gamma noise. We also showcase the good performance of our method in the case where the degradation operators are partially unknown.

Notation: We follow the notations of [22], [23]. We use lower (a) or uppercase (A) plain font for scalars, boldface lowercase ($\mathbf{a}$) for vectors, boldface uppercase ($\mathbf{A}$) for matrices and calligraphic ($\boldsymbol{\mathcal{A}}$) for tensors. The elements of vectors, matrices and tensors are denoted as a_i, $A_{i,j}$ and $\mathcal{A}_{i_1,\ldots,i_N}$, respectively. The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^T$. We use $\mathbf{I}_N$ for the $N \times N$ identity matrix and $\mathbf{0}_{L\times K}$ for the $L \times K$ matrix of zeros. Notation 1_L denotes an all-ones vector of size $L \times 1$. For a matrix $\mathbf{X}$, the notation $\mathbf{X} \geq \mathbf{0}$ means that $\mathbf{X}$ is entry-wise non-negative. Symbols $\boxtimes$ and $\odot$ denote the Kronecker and Khatri-Rao products, respectively. The Hadamard (element-wise) product is denoted by $\boxdot$. We use vec for the standard column-major vectorization of a matrix or a tensor. Each dimension of a tensor is called a mode, and the number of dimensions is called order. We restrict the scope of this paper to order-3 tensors.

II. PROBLEM FORMULATION

II-A. Prelimiaries

We introduce in Definition 2.1 the block-term decomposition with ranks $(L_r, L_r, 1)$, that we will use to build our model. The main advantage of this decomposition is to link the terms, that we will assume low-rank, to high-resolution abundance matrices and spectral signatures used in blind spectral unmixing of the unknown SRI.

Definition 2.1: **Block-term decomposition –** An order-3 tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ admits a block-term decomposition (BTD) with ranks $(L_r, L_r, 1)$ (L_rL_r1-BTD) if

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

where $\otimes$ denotes the outer product, $\mathbf{A}_r \in \mathbb{R}^{I\times L_r}$, $\mathbf{B}_r \in \mathbb{R}^{J\times L_r}$, and $\mathbf{c}_r \in \mathbb{R}^K$, for $r \in \{1,\ldots,R\}$. Moreover, we denote $\mathbf{A} = [\mathbf{A}_1,\ldots,\mathbf{A}_R] \in \mathbb{R}^{I\times\sum_r L_r}$, $\mathbf{B} = [\mathbf{B}_1,\ldots,\mathbf{B}_R] \in \mathbb{R}^{J\times\sum_r L_r}$ and $\mathbf{C} = [\mathbf{c}_1,\ldots,\mathbf{c}_R] \in \mathbb{R}^{K\times R}$.

Now, we recall in Theorem 2.2 sufficient uniqueness conditions for this decomposition[1].

Theorem 2.2 ([14]): Let $(\mathbf{A},\mathbf{B},\mathbf{C})$ denote an L_rL_r1-BTD of a tensor $\boldsymbol{\mathcal{X}}$ as in (1). Assume that $\mathbf{A}$ and $\mathbf{B}$ are full column rank and that $\mathbf{C}$ does not have proportional columns. Then $(\mathbf{A},\mathbf{B},\mathbf{C})$ is essentially unique almost surely for $r \in \{1,\ldots,R\}$.

In real-life applications such as remote sensing, these conditions are easily satisfied. Indeed, R represents the number of distinct materials in the image and we usually have $\sum_r L_r < \min(I,J,K)$. Hence the use of this decomposition for identifying more surely the ground-truth endmembers and ultimately come up with a well-grounded and powerful method to tackle our two problems.

Finally, Property 1 recalls the unfolding formulae for the L_rL_r1-BTD, that will helpful for building our algorithm:

[1]The scenario $L_r = L$ for all $r \in \{1,\ldots,R\}$ is a special case of this decomposition and is usually referred to as the LL1-BTD.

Property 1: **Tensor unfoldings –** Using the above notation, the unfoldings of a tensor $\mathcal{X}$ admitting an L_rL_r1-BTD as above can be expressed as

$$\mathbf{X}^{(1)} = \mathbf{A}\left(\mathbf{C} \odot_p \mathbf{B}\right)^T, \mathbf{X}^{(2)} = \mathbf{B}\left(\mathbf{C} \odot_p \mathbf{A}\right)^T,$$
$$\mathbf{X}^{(3)} = \mathbf{C}\left[\left(\mathbf{A}_1 \odot \mathbf{B}_1\right) 1_{L_1}, \ldots, \left(\mathbf{A}_R \odot \mathbf{B}_R\right) 1_{L_R}\right]^T$$

where $\odot_p$ denotes the partition-wise Khatri-Rao products defined as follows: $\mathbf{C} \odot_p \mathbf{A} = [\mathbf{c}_1 \boxtimes \mathbf{A}_1, \ldots, \mathbf{c}_R \boxtimes \mathbf{A}_R]$.
In Section II-B, we introduce the main assumptions, the models and associated optimization problems to build up our method for joint HSR and spectral unmixing of the unknown super-resolution image.

II-B. Assumptions, Models and Optimization problems

Let us consider two tensors $\mathcal{Y}_1 \in \mathbb{R}^{I_1 \times J_1 \times K_1}$ and $\mathcal{Y}_2 \in \mathbb{R}^{I_2 \times J_2 \times K_2}$. In the following, we assume $I_1 < I_2$, $J_1 < J_2$ et $K_2 < K_1$. In order to ease the notation, we assume that $I = I_2$, $J = J_2$ et $K = K_1$. Tensors $\mathcal{Y}_1$, $\mathcal{Y}_2$ are degraded versions of the same tensor $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$. In HSR, indices I_ℓ, J_ℓ denote the spatial dimensions whereas K_ℓ denote the spectral ones ($\ell = 1, 2$). Tensors $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively denote the HSI and MSI, whereas $\mathcal{Y}$ denotes the unknown SRI we intend to recover. We present now our main assumptions and models.

Assumption 1: **Structure of the SRI –** In the noiseless case, the tensor $\mathcal{Y}$ admits a $L_r, L_r, 1$ block-term decomposition:

$$\mathcal{Y} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^T\right) \otimes \mathbf{c}_r, \tag{2}$$

Under nonnegativity constraints, the terms $\mathbf{c}_r$ in Equation (2) can be physically interpreted as the spectral signatures associated to the R constitutive materials of $\mathcal{Y}$, while matrices $\mathbf{A}_r\mathbf{B}_r^T = \mathbf{S}_r \in \mathbb{R}^{I \times J}$ represent the corresponding abundance maps.

Assumption 2: **Structure of $\mathbf{S}_r$ –** Matrices $\mathbf{S}_r$ are assumed to be low-rank, *i.e.*,

$$\mathbf{S}_r \approx \mathbf{A}_r\mathbf{B}_r^T \in \mathbb{R}^{I \times J}, \tag{3}$$

where $\mathbf{A}_r \in \mathbb{R}^{I \times L}$ and $\mathbf{B}_r \in \mathbb{R}^{J \times L}$ admit rank L_r for all $r \in \{1, \ldots, R\}$.
The hypothesis of low-rank abundance matrices is reasonable, since the two spatial dimensions are often correlated along the rows and columns. In [24], an upper bound on the reconstruction error of such matrices by (3) is provided in the general problem. In particular, this error can be as small as desired if L_r is large enough, which motivates the Assumption 2.

Let us note $\mathbf{S} = [\text{vec}\{\mathbf{S}_1\}, \ldots, \text{vec}\{\mathbf{S}_R\}] \in \mathbb{R}^{IJ \times R}$ the matrix containing the vectorized abundance maps of each material and $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_R] \in \mathbb{R}^{L \times R}$ the matrix whose columns are the spectral signatures. The transposed third-mode unfolding of Equation (2) [12], [13] reads

$$\mathbf{Y}^{(3)T} = \mathbf{S}\mathbf{C}^T \in \mathbb{R}^{IJ \times K}, \tag{4}$$

which can be viewed as the linear mixing model (LMM) for the SRI $\mathcal{Y}$ under nonnegativity constraints. Using Assumption 2, the block-term structure (2) can thus be viewed as tensor format for the LMM, under low-rank constraints of the abundance maps.

As done in previous works (see [11]), we consider the following model providing the links between $\mathcal{Y}$ and its two degradations $\mathcal{Y}_1$ and $\mathcal{Y}_2$.

Model 1: **Tensors $\mathcal{Y}$ and ($\mathcal{Y}_1$,$\mathcal{Y}_2$) are such that**

$$\begin{cases} \mathcal{Y}_1 & \approx \sum_{r=1}^{R} \left(\mathbf{P}_1\mathbf{A}_r\left(\mathbf{P}_2\mathbf{B}_r\right)^T\right) \otimes \mathbf{c}_r, \\ \mathcal{Y}_2 & \approx \sum_{r=1}^{R} \left(\mathbf{A}_r\mathbf{B}_r^T\right) \otimes \mathbf{P}_3\mathbf{c}_r, \end{cases} \tag{5}$$

which is a coupled L_rL_r1-BTD. The tensors $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are obtained using linear downsampling operators $\mathbf{P}_1 \in \mathbb{R}^{I_1 \times I}$, $\mathbf{P}_2 \in \mathbb{R}^{J_1 \times J}$ et $\mathbf{P}_3 \in \mathbb{R}^{K_2 \times K}$, assumed to be full-rank. In a remote sensing framework, the matrix $\mathbf{P}_3 \in \mathbb{R}^{K_2 \times K}$ contains the spectral response functions for each band of the MSI sensor. The spatial degradation matrices $\mathbf{P}_1 \in \mathbb{R}^{I_1 \times I}$ and $\mathbf{P}_2 \in \mathbb{R}^{J_1 \times J}$ perform Gaussian blurring and downsampling along each spatial dimension, *i.e.* we suppose that the spatial degradation operation is separable, as in the commonly used Wald's protocol [25]. The approximately equal symbols in Equation (5) account for the presence of noise during the degradation process.

State-of-the-art unmixing algorithms aim at recovering $\{\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^T\}_{r=1}^R$ and $\mathbf{C}$ from the mixed pixels in $\mathcal{Y}$. Here, since $\mathcal{Y}$ is unknown and only $\mathcal{Y}_1$ is observed with high spectral resolution, these algorithms are only able to recover spatially-degraded versions of the abundance maps [13], namely

$$\mathbf{P}_1\mathbf{S}_r\mathbf{P}_2^T \in \mathbb{R}^{I_H \times J_H} \text{ for } r \in \{1, \ldots, R\}. \tag{6}$$

Differently from those works, fusion of an HSI with an MSI with high spatial resolution allows us to seek for abundance maps at a higher spatial resolution.

Thus jointly solving the data fusion and blind unmixing problems consists in finding the LL1 factors $\{\mathbf{A}_r\mathbf{B}_r^T\}_{r=1}^R$, $\mathbf{C}$, under the assumption of (5), subject to the constraints

$$\{\mathbf{A}_r\mathbf{B}_r^T\}_{r=1}^R \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0}, \mathbf{P}_i \geq \mathbf{0} \text{ for } i \in \{1, \ldots, 3\}. \tag{7}$$

Trying to minimize the approximation errors in (5) leads, for instance, to minimizing the following cost function:

$$\begin{aligned} \Phi = & D_\beta\left(\mathcal{Y}_1 \| \sum_{r=1}^{R} \left(\mathbf{P}_1\mathbf{A}_r\left(\mathbf{P}_2\mathbf{B}_r\right)^T\right) \otimes \mathbf{c}_r\right) \\ & + \lambda D_\beta\left(\mathcal{Y}_2 \| \sum_{r=1}^{R} \left(\mathbf{A}_r\mathbf{B}_r^T\right) \otimes \mathbf{P}_3\mathbf{c}_r\right), \end{aligned} \tag{8}$$

where λ is a positive penalty parameter, subject to the constraints in (7). For a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$,

$$D_\beta\left(\boldsymbol{\mathcal{Y}}\|\sum_{r=1}^{R}\left(\mathbf{A}_r\mathbf{B}_r^T\right)\otimes\mathbf{c}_r\right) = \sum_{i,j,k} d_\beta\left((\mathcal{Y})_{i,j,k}\|\left((\mathbf{A}_r)_{i,:}(\mathbf{B}_r)_{:,j}^T\right)\otimes(\mathbf{c}_r)_k\right), \tag{9}$$

with $d_\beta(x\|y)$ the β-divergence between the two scalars x and y. For $\beta = 2$, this amounts to the standard squared Euclidean distance since $d_2(x\|y) = \frac{1}{2}(x-y)^2$. For $\beta = 1$ and $\beta = 0$, the β-divergence corresponds to the Kullback-Leibler (KL) divergence and the Itakura-Saito (IS) divergence, respectively. For Nonnegative Matrix Factorization (NMF) models, the data fitting term should be chosen depending on the noise statistic assumed in the generative model of the data, see [26], [27], [28], [21] and references therein for more details. In Section II-C, we present our Algorithm to tackle the family of optimization problems given in Equation (8).

II-C. Algorithms

Most nonnegative tensor decomposition algorithms are based on an iterative scheme that alternatively update one factor at the time with the others kept fixed, and we adopt this approach in this paper. The goal in this section is to derive an algorithm to solve (8) based on the multiplicative updates (MU). Let us consider the subproblem in $\mathbf{A}$ (with the others fixed) after unfolding along the first mode following Property 1:

$$\min_{A\geq 0} D_\beta(\mathbf{Y}_1^{(1)}\|\mathbf{P}_1\mathbf{A}(\mathbf{C}\odot_p\mathbf{P}_2\mathbf{B})^T) + \lambda D_\beta(\mathbf{Y}_2^{(1)}\|\mathbf{A}(\mathbf{P}_3\mathbf{C}\odot_p\mathbf{B})^T). \tag{10}$$

To tackle this problem, we follow the standard majorization-minimization (MM) framework [29] and the results given by [21, Lemma 2]. Given the current iterate $\tilde{\mathbf{A}}$, let us pose $\mathbf{H}_1 = (\mathbf{C}\odot_p\mathbf{P}_2\mathbf{B})^T$ and $\mathbf{H}_2 = (\mathbf{P}_3\mathbf{C}\odot_p\mathbf{B})^T$, we obtain the following update:

$$\mathbf{A} = \tilde{\mathbf{A}}\boxdot\left(\frac{\left[\mathbf{P}_1^T\left(\left(\mathbf{P}_1\tilde{\mathbf{A}}\mathbf{H}_1\right)^{\cdot(\beta-2)}\boxdot\mathbf{Y}_1^{(1)}\right)\mathbf{H}_1^T + \lambda\left(\left(\tilde{\mathbf{A}}\mathbf{H}_2\right)^{\cdot(\beta-2)}\boxdot\mathbf{Y}_2^{(1)}\right)\mathbf{H}_2^T\right]}{\left[\mathbf{P}_1^T\left(\mathbf{P}_1\tilde{\mathbf{A}}\mathbf{H}_1\right)^{\cdot(\beta-1)}\mathbf{H}_1^T + \lambda\left(\tilde{\mathbf{A}}\mathbf{H}_2\right)^{\cdot(\beta-1)}\mathbf{H}_2^T\right]}\right)^{\cdot\gamma(\beta)} \tag{11}$$

where $A\boxdot B$ (resp. $\frac{[A]}{[B]}$) is the Hadamard product (resp. division) between A and B, $A^{(\cdot\alpha)}$ is the element-wise α exponent of A, $\gamma(\beta) = \frac{1}{2-\beta}$ for $\beta < 1$, $\gamma(\beta) = 1$ for $\beta\in[1,2]$ and $\gamma(\beta) = \frac{1}{\beta-1}$ for $\beta > 2$ [30]. The subproblems in $\mathbf{B}$ and $\mathbf{C}$ can be solved similarly and their closed form expressions can be found in Appendix A.

Contrary to the majority of state-of-the-art methods, our algorithms are also able to estimate the degradation matrices $\mathbf{P}_i$ for $i\in\{1,...,3\}$. These updates can be derived based on the classical MU associated to the matrix model $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ [30]. For $\mathbf{P}_1$, we are interested in solving $D_\beta(\mathbf{Y}_1^{(1)}\|\mathbf{P}_1\mathbf{A}(\mathbf{C}\odot_p\mathbf{P}_2\mathbf{B})^T)$. By posing $\mathbf{V}^T = \mathbf{A}(\mathbf{C}\odot_p\mathbf{P}_2\mathbf{B})^T$, we derive:

$$\mathbf{P}_1 \leftarrow \widetilde{\mathbf{P}}_1\boxdot\left(\frac{\left[\left(\widetilde{\mathbf{P}}_1\mathbf{V}^T\right)^{\cdot(\beta-2)}\boxdot\mathbf{Y}_1^{(1)}\right]\mathbf{V}}{\left[\widetilde{\mathbf{P}}_1\mathbf{V}^T\right]^{\cdot(\beta-1)}\mathbf{V}}\right)^{\cdot\gamma(\beta)}. \tag{12}$$

Similar rationale has been followed for the updates of $\mathbf{P}_2$ and $\mathbf{P}_3$, see Appendix B for more details.

Algorithm 1 summarizes our method to tackle (8) which will be referred to as β-(L_r,L_r,1)-NBTD. It consists in two optimization loops:

Loop 1: $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ only are alternatively updated with downsampling matrices fixed for a maximum of `i1` iterations. $\mathbf{P}_i$ for $i\in\{1,...,3\}$ kept fixed to obtain good estimates for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$.

Loop 2: All the factors, including the matrices $\mathbf{P}_i$, are alternatively updated. The maximum number of iterations for Loop 2 is `i2`. For the HSR problem, the operators $\mathbf{P}_i$ for $i\in\{1,...,3\}$ are usually known and therefore the parameter `i2` is set to zero. Loop 2 is considered in the case we have partial knowledge or uncertainties on one of more downsampling operators, similarly as done in [21] with a matrix model. This case will be later referred to as "semi-blind".

The Algorithm is stopped when the relative change of the cost function Φ from (8) is below some given threshold κ, and when the maximum number of iterations is reached.

Initialization: Many options are available to initialize $(\mathbf{A},\mathbf{B},\mathbf{C})$. Traditionally, MU-based methods are initialized with random factors. In this work, we propose an efficient way to initialize the low-rank factors of the L_rL_r1-BTD. For $\mathbf{C}$, one can draw independent and identically distributed entries from absolute continuous joint distributions. In our experiments, we initialized $\mathbf{C}$ by performing VCA [31] on the HSI $\boldsymbol{\mathcal{Y}}_1$, thus extracting high-resolution spectral information. Then, the matrix $\mathbf{S}$ of vectorized abundance maps is obtained by solving the following inverse problem:

$$\mathbf{S}^T = (\mathbf{P}_3\mathbf{C})^\dagger\,\mathbf{Y}_2^{(3)}. \tag{13}$$

Initialization of $\mathbf{A}_r$ and $\mathbf{B}_r$ for $r\in\{1,\ldots,R\}$ is performed based on the classical MU [30] with a maximum of `j1` iterations, yielding the following updates:

$$\begin{aligned}\mathbf{A}_r &\leftarrow \widetilde{\mathbf{A}}_r\boxdot\left(\frac{\left[\left(\widetilde{\mathbf{A}}_r\mathbf{B}_r^T\right)^{\cdot(\beta-2)}\boxdot\mathbf{S}_r\right]\mathbf{B}_r}{\left[\widetilde{\mathbf{A}}_r\mathbf{B}_r^T\right]^{\cdot(\beta-1)}\mathbf{B}_r}\right)^{\cdot\gamma(\beta)},\\ \mathbf{B}_r &\leftarrow \widetilde{\mathbf{B}}_r\boxdot\left(\frac{\mathbf{A}_r^T\left[\left(\mathbf{A}_r\widetilde{\mathbf{B}}_r^T\right)^{\cdot(\beta-2)}\boxdot\mathbf{S}_r\right]}{\mathbf{A}_r^T\left[\mathbf{A}_r\widetilde{\mathbf{B}}_r^T\right]^{\cdot(\beta-1)}}\right)^{\cdot\gamma(\beta)}.\end{aligned} \tag{14}$$

In the blind case, *i.e.*, when one or several matrices $\mathbf{P}_i$ are unknown, they are initialized similarly using (12) with a maximum of `j2` iterations. The initialization procedure is summarized in Algorithm 2.

Algorithm 1 MU for β-(L_r,L_r,1)-NBTD

Input: Input nonnnegative tensors $\mathcal{Y}_1$ and $\mathcal{Y}_2$, nonnegative initializations $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, nonnegative downsampling operators $\mathbf{P}_i$ for $i \in \{1, ..., 3\}$; R, ranks $\{L_r\}_{r=1}^R$, maximum number of iterations `i1` and `i2`, a threshold $0 < \kappa \ll 1$, and a weight $\lambda > 0$.

Output: An approximate solution to (8) under constraints (7)

% Loop 1
$i \leftarrow 0$, $\Phi^0 = 1$, $\Phi^1 = 0$.
while $i <$ `i1` and $\left|\frac{\Phi^i - \Phi^{i+1}}{\Phi^i}\right| > \kappa$ **do**
    % Update of matrices $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$
    Update $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ sequentially; see Equations (11)
    Compute the objective function Φ^{i+1}
end while
% Loop 2
$i \leftarrow 0$
while $i <$ `i2` and $\left|\frac{\Phi^i - \Phi^{i+1}}{\Phi^i}\right| > \kappa$ **do**
    % Update of factors
    Update $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{P}_i$ for $i \in \{1, ..., 3\}$ sequentially; see Equations (11) and (12)
    Compute the objective function Φ^{i+1}
end while
return $\hat{\mathcal{Y}} = \sum_{r=1}^R \left(\mathbf{A}_r \mathbf{B}_r^T\right) \otimes \mathbf{c}_r$

Algorithm 2 Initialization of Algorithm 1

Input: Input nonnnegative tensors $\mathcal{Y}_1$ and $\mathcal{Y}_2$, nonnegative downsampling operators $\mathbf{P}_i$ for $i \in \{1, ..., 3\}$; R, ranks $\{L_r\}_{r=1}^R$, maximum number of iterations `j1` and `j2`, a threshold $0 < \kappa \ll 1$.

Output: Initial values $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$

Initialize $\mathbf{C}$ using VCA on $\mathcal{Y}_1$;
Compute $\mathbf{S}$ using (13);
% Non-blind case: update of $\mathbf{A}_r$, $\mathbf{B}_r$ *only*
$j \leftarrow 0$, $\Phi^0 = 1$, $\Phi^1 = 0$.
while $j <$ `j1` and $\left|\frac{\Phi^j - \Phi^{j+1}}{\Phi^j}\right| > \kappa$ **do**
    Update $\mathbf{A}_r$, $\mathbf{B}_r$, see (14);
    Compute the objective function Φ^{j+1}
end while
% Blind case: Update $\mathbf{A}_r$, $\mathbf{B}_r$, *and* $\mathbf{P}_i$ *for* $i \in \{1, \ldots, 3\}$
$j \leftarrow 0$
while $j <$ `j2` and $\left|\frac{\Phi^j - \Phi^{j+1}}{\Phi^j}\right| > \kappa$ **do**
    % Update of factors
    Update $\mathbf{A}_r$, $\mathbf{B}_r$, and $\mathbf{P}_i$; see Equations (14) and (12)
    Compute the objective function Φ^{j+1}
end while
return $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_R]$, $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_R]$ and $\mathbf{C}$

Choice of the ranks: Regarding R, intuitively one would select the real number of materials in the image. In real applications, reference mixing factors are unknown, thus R is unknown as well. Therefore, in the absence of reference endmembers, R is estimated based on e.g., subspace identification [32], [33] Then, the L_r are chosen to be large while satisfying the conditions of Theorem 2.2. Several works considered mixed-norm regularization to estimate the L_r for various LL1-BTD, but these methods were not guaranteed to provide a unique solution when applied to the unmixing problem [13], [15]. Conversely, in [14] were given guarantees for unique recovery of the SRI and its mixing factors.

Comments on computational complexity: It can be verified that the computational complexity of the β-(L_r,L_r,1)-NBTD is asymptotically equivalent to the standard MU for β-NMF (after modal unfoldings), that is, it requires $\mathcal{O}(IJK \times \sum_r^R L_r)$ operations per iteration.

Parallelization: Some of the most computationally intensive steps of the proposed algorithm can be easily ran onto a parallel computation platform. Indeed, the complexity of our MU given in Equation (11), for instance, is mainly driven by the matrix products in which matrices $\mathbf{A}$ and $\mathbf{H}$ are involved. On Matlab for example, one can easily take of advantage of a GPU compatible with CUDA libraries by simply transforming usual arrays into GPU arrays and significantly speed up the Algorithm.

Comments on convergence guarantees: In practice, for the updates of factors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{P}_i$, we take the element-wise maximum between the matrix updates, that correspond to the closed form expression of the minimizer of the majorization built at the current iterate [21], [30], and a small positive scalar ϵ (here we choose the Matlab machine epsilon). These modified updates aim at establishing convergence guarantee to stationary points within the Block Successive Minimization Methods (BSUM) framework [34].

III. NUMERICAL EXPERIMENTS

We now apply our method on synthetic data sets. All tests are preformed using Matlab R2021a on a laptop Intel CORE i7-11800H CPU @2.30GHz 16GB RAM with GeForce RTX3060 GPU. The code is available from "insert link here".

III-A. Test setup

Our method was tested against several tensorial methods recently and successfully used for solving the HSR problem, namely STEREO and Blind-STEREO [35], [36], SCOTT and BSCOTT [37], CT-STAR and CB-STAR [38], and SCLL1 [13] and CNN-BTD-Var [14]. Among them, SCLL1 and CNN-BTD-Var were based on the LL1-BTD model, therefore they are able to solve the unmixing problem[2]. We

[2]The performance of SCLL1 for blind spectral unmixing was not assessed in the original work.

also benchmarked several matrix-based approaches: CNMF [3], FUSE [2] , HySure [5] and SFIM [2]. Being based on coupled nonnegative matrix factorization, CNMF was able to perform joint fusion and unmixing. We chose the ranks and regularization parameters for these algorithms according to original works.

We initialized our algorithm using Algorithm 2 with a maximum of 500 iterations and a threshold $\kappa = 10^{-7}$. We ran our method with a maximum of 1000 iterations for each loop and $\kappa = 10^{-7}$.

We compared the groundtruth SRI $\mathcal{Y}$ with the estimated SRI $\widehat{\mathcal{Y}}$ obtained by the algorithms. The main performance metric used in comparisons was the *Peak Signal-to-Noise ratio* (PSNR) [39]:

$$\text{PSNR} = 10\log_{10}\left(\frac{\|\mathcal{Y}\|_F^2}{\|\widehat{\mathcal{Y}} - \mathcal{Y}\|_F^2}\right). \tag{15}$$

In addition to PSNR, we considered different metrics [39] described below:

$$\text{CC} = \frac{1}{IJK}\left(\sum_{k=1}^{K}\rho\left(\mathcal{Y}_{:,:,k}, \widehat{\mathcal{Y}}_{:,:,k}\right)\right), \tag{16}$$

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

$$\text{ERGAS} = \frac{100}{d}\sqrt{\frac{1}{IJK}\sum_{k=1}^{K}\frac{\|\widehat{\mathcal{Y}}_{:,:,k} - \mathcal{Y}_{:,:,k}\|_F^2}{\mu_k^2}}, \tag{17}$$

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

$$\text{SAD} = \frac{1}{R}\sum_{r=1}^{R}\arccos\left(\frac{\mathbf{c}_r^T\widehat{\mathbf{c}}_r}{\|\mathbf{c}_r\|_2\|\widehat{\mathbf{c}}_r\|_2}\right), \tag{18}$$

which computes the spectral angle distance between original and estimated spectra, and can be used to assess unmixing performance as well. Performance for recovery of the abundance maps was assessed using the root mean-squared error between reference $\mathbf{S}$ and estimate $\widehat{\mathbf{S}}$:

$$\text{RMSE} = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{IJ}\sum_{d=1}^{IJ}\left((\mathbf{S}_r)_d - (\widehat{\mathbf{S}}_r)_d\right)\Big)^2}. \tag{19}$$

Finally, we considered the computational time for each algorithm, given by the `tic` and `toc` functions of Matlab.

III-B. Datasets

Degradation model: In our experiments, we considered two different datasets for which a reference SRI was available. The HSI was obtained by spatial degradation of $\mathcal{Y}$ by $\mathbf{P}_1$ and $\mathbf{P}_2$ while the MSI was obtained by spectral degradation of $\mathcal{Y}$ by $\mathbf{P}_3$ according to model (5).

For spatial degradation, we followed the commonly used Wald's protocol [25]. The matrices $\mathbf{P}_1$, $\mathbf{P}_2$ were computed with a separable Gaussian blurring kernel of size q. Downsampling was performed along each spatial dimension with a ratio d between the SRI and HSI, as in previous works [8]–[13]. Refer to Appendix B for more details on the construction of $\mathbf{P}_1$, $\mathbf{P}_2$.

For the spectral degradation matrix $\mathbf{P}_3$, we used the spectral response functions of the Sentinel-2 instrument[3]. It spans the electromagnetic spectrum from 412nm to 2022nm and produced a 10-band MSI corresponding to the wavelengths 433–453nm (atmospheric correction), 458–522nm (soil, vegetation), 543–577nm (green peak), 650–680nm (maximum chlorophyll absorption), 698–712nm (red edge), 733–747nm (red edge), 773–793nm (leaf area index, edge of NIR), 785–900nm (leaf area index), 855–875nm (NIR plateau), 935–955nm (water vapour absorption). The spectral degradation matrix $\mathbf{P}_3$ was a selection-weighting matrix that selected the common spectral bands of the SRI and the MSI.

As done in [21], three noise statistics were considered that are; Gaussian Noise, Poisson noise and Multiplicative Gamma noise. For the first two, a SNR of 30dB has been considered while for Gamma noise, we considered a distribution of mean 1 and variance 0.05. Therefore, in our experiments, we had $\lambda = 1$.

Synthetic dataset: We first assessed the unmixing performance in the case where the SRI admits an exact L_rL_r1-BTD. Although this dataset did real spectral images, it allowed us to assess unmixing performance in a case where the uniqueness conditions for the non-negative matrix factorization model (see [40], [41]) were not fulfilled.

We considered $R = 4$ spectral signatures $\mathbf{c}_r$ ($r \in \{1,\ldots,R\}$) obtained from the Jasper Ridge reference data[4]. The SRI $\mathcal{Y} \in \mathbb{R}^{I\times J\times K}$ ($I = J = 120$, $K = 173$) was split into 36 equal blocks in the spatial dimensions. We set $L_1 = L_2 = 3$ and $L_3 = L_4 = 6$. Each abundance map $\mathbf{S}_r$ ($r \in \{1,\ldots,R\}$) was a block matrix with L_r blocks of size $\frac{I}{L_r}\times\frac{J}{L_r}$ such that the pure pixel assumption was valid. Thus in each block, at most one material was active, as indicated by the parcel map shown in Figure 1. Each block in the parcel map was a patch composed of entries equal to one. The abundance maps resembled agricultural fields. This was a case for which unconstrained non-negative matrix factorization was not unique.

[3]available for download at https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/assetpublisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses.

[4]Available for download at http://lesun.weebly.com/hyperspectral-data-set.html.

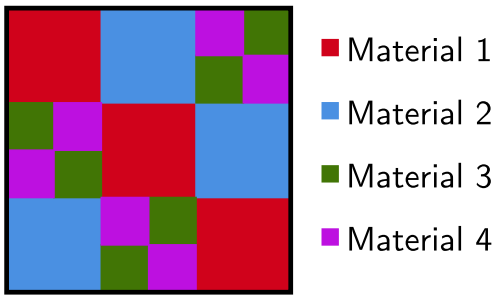


Fig. 1. Parcel map for the first dataset

Formally, we computed the reference SRI as

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \mathbf{S}_r \otimes \mathbf{c}_r.$$

For $\mathbf{P}_1 = \mathbf{P}_2$, we had $q = 9$ and $d = 4$ so that $I_H = J_H = 30$. For $\mathbf{P}_3$, the spectral response of the Sentinel-2 MS sensor led to $K_M = 10$.

Semi-real datasets: Second, we considered a semi-real dataset based on the Jasper Ridge reference SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{100\times100\times173}$. This dataset was composed of four materials: road, soil, water and vegetation. For $\mathbf{P}_1 = \mathbf{P}_2$, we had $q = 9$ and $d = 4$, and $K_M = 10$. We chose $R = 4$, and $L_1 = 15$, $L_2 = 8$, $L_3 = 20$, $L_4 = 13$.

We also considered a semi-real dataset based on the Samson reference SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{95\times95\times156}$. This dataset was composed of three materials: soil, water and vegetation. For $\mathbf{P}_1 = \mathbf{P}_2$, we had $q = 9$ and $d = 5$, and $K_M = 10$. We chose $R = 3$, and $L_1 = 15$, $L_2 = 12$, $L_3 = 10$.

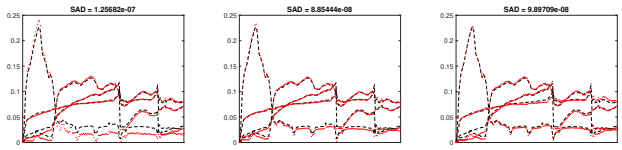
III-C. Results

In order to assess the performance of our approach for data fusion, we reported the quality metrics obtained with each method over 5 trials in the tables below. The two best metrics of each columns were shown in bold.

For the unmixing task, we plotted the reference and estimated spectral signatures and abundance maps obtained by our algorithm. We computed the SAD and RMSE on the permuted materials so that they match the reference.

Results on synthetic dataset:

Table I reports the quality metrics obtained for the synthetic dataset. Except for runtimes, one can observe that our method competes favorably with SOTA methods, in particular for Gaussian and Gamma noise settings, our method respectively ranks third and second. Among the LL1-based methods, our algorithm provided the best results, thus highlighting the gain in flexibility provided by the use of L_r different from each other. Other benchmarked tensor methods also yielded high performance, but they were not suited for spectral unmixing. Figures 2 and 3 show the unmixing results obtained with our method, depicting a correct estimation of ground-truth spectral signatures and abundance maps for the various noise statistics.



(a) (b) (c)

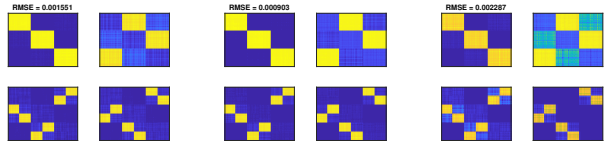
Fig. 2. Reference (black dashed line) and estimated (red dots) spectral signatures with (a) Gaussian noise, (b) Poisson noise, (c) Gamma noise, synthetic dataset.

Method	CC	SAD	RMSE	ERGAS	PSNR (dB)	Time (sec)
Best	1	0	0	0	∞	0
Gaussian Noise - 30dB						
STEREO	0.997	4.43	9.339e-3	2.699	34.44	1.452
BSTEREO	0.990	8.37	1.876e-2	4.473	29.19	1.338
SCOTT	0.993	6.57	1.614e-2	3.310	32.04	2.047
BSCOTT	0.994	5.40	1.382e-2	3.419	32.11	0.228
SCLL1	0.988	5.79	3.297e-2	5.148	25.15	34.13
CT-STAR	**0.999**	**0.51**	**1.332e-3**	**0.919**	**38.92**	**0.128**
CB-STAR	**0.999**	2.76	**6.399e-3**	**1.578**	**38.17**	16.36
BTD-Var	0.996	5.33	1.270e-2	2.980	33.75	3.645
CNMF	0.999	2.57	1.238e-2	2.113	31.68	1.930
FUSE	0.995	3.79	1.341e-2	2.826	31.64	**0.148**
HySure	0.990	8.17	2.025e-2	4.114	28.61	16.75
SFIM	0.981	9.77	3.455e-2	5.384	25.92	0.254
Alg. 1 ($\beta = 2$)	0.997	**2.16**	1.043e-2	2.240	37.72	35.22
Poisson Noise - 30dB						
STEREO	0.999	2.82	8.882e-3	1.347	35.31	1.573
BSTEREO	0.995	7.98	2.134e-2	2.809	29.93	1.379
SCOTT	0.999	2.77	1.122e-2	1.518	34.30	2.087
BSCOTT	0.999	1.47	1.146e-2	1.601	34.11	0.204
SCLL1	0.987	5.70	3.609e-2	5.434	24.46	33.57
CT-STAR	**0.999**	**0.25**	**5.272e-3**	**0.713**	**39.71**	**0.102**
CB-STAR	**0.999**	1.25	**6.408e-3**	0.991	**38.02**	20.14
CNN-BTD-Var	0.997	5.30	1.260e-2	2.952	32.76	3.302
CNMF	0.999	**1.09**	7.313e-3	1.364	36.20	1.872
FUSE	0.994	2.77	1.470e-2	2.813	30.92	**0.146**
HySure	0.998	3.04	1.375e-2	2.174	32.35	17.48
SFIM	0.990	4.96	2.544e-2	3.817	28.07	0.281
Alg. 1 ($\beta = 1$)	0.999	1.19	1.065e-2	**0.776**	35.69	23.42
Gamma Noise						
STEREO	0.996	7.03	1.645e-2	2.249	31.36	1.986
BSTEREO	0.991	11.1	2.739e-2	3.355	28.60	1.677
SCOTT	0.991	8.45	2.935e-2	3.578	31.56	3.306
BSCOTT	0.988	3.88	3.127e-2	4.004	29.69	0.385
SCLL1	0.980	5.76	3.357e-2	6.066	24.58	42.711
CT-STAR	**0.999**	**0.60**	**2.700e-3**	**0.501**	**37.79**	**0.160**
CB-STAR	0.999	2.44	**7.975e-3**	**1.356**	37.31	27.326
CNN-BTD-Var	0.995	5.39	1.338e-2	3.357	32.85	4.909
CNMF	0.997	1.31	1.591e-2	2.501	30.33	2.060
FUSE	0.993	3.11	1.690e-2	3.139	29.93	**0.175**
HySure	0.988	5.96	3.253e-2	4.324	28.37	23.871
SFIM	0.986	5.82	3.195e-2	4.660	26.32	0.368
Alg. 1 ($\beta = 0$)	**0.999**	**0.72**	8.487e-3	9.736	**37.90**	64.243

Table I. Reconstruction metrics, synthetic dataset.

Results on semi-real dataset:

Table II reports the quality metrics obtained for the semi-real dataset. Although amongst the slowest methods, our algorithm performed relatively well on this dataset. In particular, it provided the best results amongst the LL1-based approaches with Gaussian noise and had performance comparable to that of most approaches. It ranked generally third best for Poisson noise, and yielded the best metrics



(a) (b) (c)

Fig. 3. Reference (black dashed line) and estimated (red dots) abundance maps with (a) Gaussian noise, (b) Poisson noise, (c) Gamma noise, synthetic dataset.

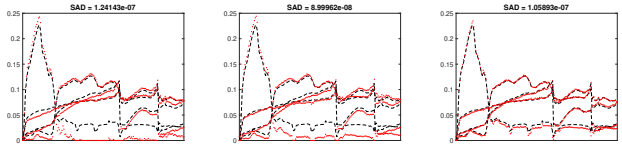
for Gamma noise. In Figures 4 and 5, one can see the high quality of the unmixing results.

Method Best	CC 1	SAD 0	RMSE 0	ERGAS 0	PSNR (dB) ∞	Time (sec) 0
Gaussian Noise - 30dB						
STEREO	0.987	3.19	1.272e-2	2.629	**31.59**	1.647
BSTEREO	0.987	3.25	**1.299e-2**	2.598	**31.39**	1.407
SCOTT	0.982	4.60	1.732e-2	3.250	29.52	**0.170**
BSCOTT	0.973	4.67	2.078e-2	4.041	28.34	0.197
SCLL1	0.976	5.78	2.957e-2	3.713	25.62	11.98
CT-STAR	0.886	8.66	5.304e-2	6.876	19.43	**0.067**
CB-STAR	**0.990**	3.16	1.296e-2	**2.240**	31.39	10.59
CNN-BTD-Var	0.920	8.44	4.084e-2	5.677	21.47	1.267
CNMF	**0.998**	**1.59**	**9.373e-3**	**1.587**	33.83	1.338
FUSE	0.983	3.47	1.962e-2	2.933	28.35	0.242
HySure	0.980	6.77	1.964e-2	3.941	28.62	15.08
SFIM	0.975	7.77	2.911e-2	4.013	27.41	0.445
Alg. 1 ($\beta = 2$)	0.987	**3.11**	1.849e-2	2.640	28.79	43.54
Poisson Noise - 30dB						
STEREO	0.989	3.01	1.378e-2	2.617	**31.10**	1.597
BSTEREO	0.989	3.05	1.435e-2	2.479	30.69	1.673
SCOTT	0.991	3.30	1.584e-2	2.361	30.20	**0.180**
BSCOTT	0.981	3.55	2.282e-2	3.191	28.11	0.213
SCLL1	0.979	5.23	2.983e-2	3.656	25.46	11.387
CT-STAR	0.886	8.63	5.328e-2	7.204	19.53	**0.092**
CB-STAR	0.992	2.89	1.371e-2	2.269	31.09	7.592
CNN-BTD-Var	0.920	8.63	4.152e-2	6.078	21.45	1.130
CNMF	**0.999**	**0.90**	**6.000e-3**	**1.493**	27.96	1.302
FUSE	0.984	2.88	2.066e-2	3.227	28.15	0.276
HySure	**0.998**	**2.05**	**1.196e-2**	**1.778**	**33.55**	13.734
SFIM	0.989	3.71	2.083e-2	2.832	29.54	0.320
Alg. 1 ($\beta = 1$)	0.990	3.16	1.712e-2	2.657	29.58	42.241
Gamma Noise						
STEREO	0.986	3.83	1.667e-2	2.973	29.60	1.549
BSTEREO	0.986	4.01	1.720e-2	2.856	29.38	1.391
SCOTT	0.966	8.19	4.495e-2	4.815	27.46	0.147
BSCOTT	0.967	5.62	3.741e-2	4.192	26.32	**0.163**
SCLL1	0.969	5.77	3.706e-2	4.410	24.17	10.579
CT-STAR	0.885	8.66	5.308e-2	7.292	19.46	**0.070**
CB-STAR	**0.990**	3.23	**1.475e-2**	**2.547**	**30.53**	8.243
CNN-BTD-Var	0.917	8.75	4.181e-2	6.241	21.36	1.005
CNMF	0.970	3.57	3.095e-2	4.316	26.20	0.960
FUSE	0.981	**3.09**	2.191e-2	3.432	27.64	0.217
HySure	0.986	5.01	2.968e-2	3.435	28.99	11.662
SFIM	0.983	4.56	2.764e-2	3.341	27.27	0.299
Alg. 1 ($\beta = 0$)	**0.996**	**1.36**	**1.389e-2**	**2.058**	**31.66**	59.298

Table II. Reconstruction metrics, semi-real dataset.

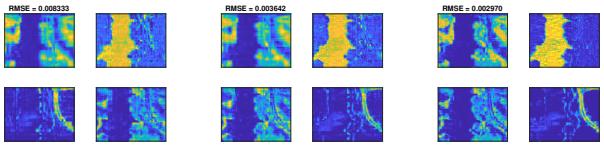
Results on semi-real dataset with estimation of $\mathbf{P}_1$ and $\mathbf{P}_2$:

We now assess the performance of our method in the semi-blind case. In this part, we consider the Samson-based dataset and we consider $\mathbf{P}_1$ and $\mathbf{P}_2$ to be unknown, while $\mathbf{P}_3$ is known, with Gaussian noise. We compare our algorithm with `i2` > 0 to two semi-blind tensor approaches, respectively BSTEREO and BSCOTT. In Table III, we can see that except for PSNR, our algorithm yielded the best metrics. In



(a) (b) (c)

Fig. 4. Reference (black dashed line) and estimated (red dots) spectral signatures with (a) Gaussian noise, (b) Poisson noise, (c) Gamma noise, semi-real dataset.



(a) (b) (c)

Fig. 5. Reference (black dashed line) and estimated (red dots) abundance maps with (a) Gaussian noise, (b) Poisson noise, (c) Gamma noise, semi-real dataset.

particular, its SAD and RMSE were way lower than those of the benchmarked algorithms, since our approach is the only one able to perform unmixing. Our algorithm was also the only one to be able to estimate the degradation matrices. In Figure 6, we could see the correct reconstruction of the spectral signatures. Finally, we computed the relative error between the reference HSI $\mathcal{Y}_1$ and the tensor obtained from degradation of the SRI with the estimated degradation matrices, namely $\widehat{\mathcal{Y}}_1 = \mathcal{Y} \bullet_1 \widehat{\mathbf{P}}_1 \bullet_2 \widehat{\mathbf{P}}_2$. Our approach yielded an error of $1.6 \cdot 10^{-3}$.

Method Best	CC 1	SAD 0	RMSE 0	ERGAS 0	PSNR (dB) ∞	Time (sec) 0
Gaussian Noise - 30dB						
Alg. 1 ($\beta = 2$)	**0.980**	**3.56**	**1.930e-2**	**2.704**	27.74	22.315
BSTEREO	0.975	6.84	2.514e-2	3.188	25.72	1.755
BSCOTT [4,4] blocks	0.967	5.94	2.061e-2	3.506	**29.03**	**0.050**

Table III. Reconstruction metrics, semi-real dataset, blind case.

IV. CONCLUSIONS

In this paper, we have shown that nonnegative block-term decompositions along with coupled optimization problems using β-divergence can be used meaningfully for jointly solving the super-resolution and blind spectral unmixing problems. We have provided a family of simples algorithm to tackle these problems and have illustrated the behaviour of the method on synthetic and semi-real datasets with various noise statistics. Further works will focus on the development of penalized variants of Problems (8). In particular, we will investigate the use of minimum-volume constraints as a rank-estimation strategy. Efficient optimization algorithms will

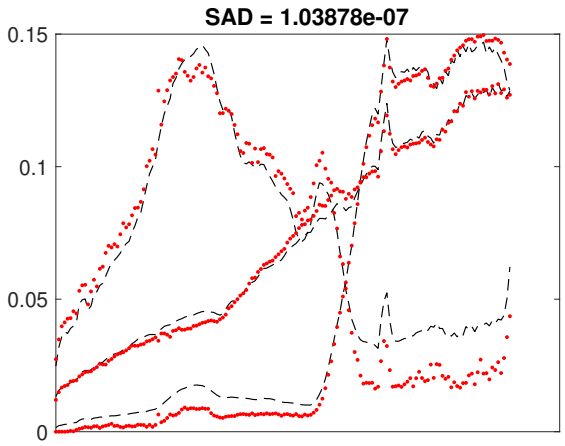


Fig. 6. Reference (black dashed line) and estimated (red dots) spectral signatures with Gaussian noise, semi-real dataset, blind case.

also be envisioned. A parallelized version of our algorithm will highly decrease its computation time. Extended test cases, including almost collinear materials and/or inter-image variability, will be addressed.

Appendix A
Detailed MU updates

In this appendix, we give the closed-form expressions for the multiplicative updates of $\mathbf{B}$ and $\mathbf{C}$, similarly to (11).

$$\mathbf{B} = \tilde{\mathbf{B}} \boxdot \left(\frac{\left[\mathbf{P}_2^T \left(\left(\mathbf{P}_2 \tilde{\mathbf{B}} \mathbf{H}_1 \right)^{\cdot(\beta-2)} \boxdot \mathbf{Y}_1^{(2)} \right) \mathbf{H}_1^T + \lambda \left(\left(\tilde{\mathbf{B}} \mathbf{H}_2 \right)^{\cdot(\beta-2)} \boxdot \mathbf{Y}_2^{(2)} \right) \mathbf{H}_2^T \right]}{\left[\mathbf{P}_2^T \left(\mathbf{P}_2 \tilde{\mathbf{B}} \mathbf{H}_1 \right)^{\cdot(\beta-1)} \mathbf{H}_1^T + \lambda \left(\tilde{\mathbf{B}} \mathbf{H}_2 \right)^{\cdot(\beta-1)} \mathbf{H}_2^T \right]} \right)^{\cdot\gamma(\beta)}, \tag{20}$$

where $\mathbf{H}_1 = (\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A})^T$ and $\mathbf{H}_2 = (\mathbf{P}_3 \mathbf{C} \odot_p \mathbf{A})^T$. For $\mathbf{C}$, we have

$$\mathbf{C} = \tilde{\mathbf{C}} \boxdot \left(\frac{\left[\left(\left(\tilde{\mathbf{C}} \mathbf{H}_1 \right)^{\cdot(\beta-2)} \boxdot \mathbf{Y}_1^{(3)} \right) \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \left(\left(\mathbf{P}_3 \tilde{\mathbf{C}} \mathbf{H}_2 \right)^{\cdot(\beta-2)} \boxdot \mathbf{Y}_2^{(3)} \right) \mathbf{H}_2^T \right]}{\left[\left(\tilde{\mathbf{C}} \mathbf{H}_1 \right)^{\cdot(\beta-1)} \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \left(\mathbf{P}_3 \tilde{\mathbf{C}} \mathbf{H}_2 \right)^{\cdot(\beta-1)} \mathbf{H}_2^T \right]} \right)^{\cdot\gamma(\beta)}, \tag{21}$$

with $\mathbf{H}_1 = (\mathbf{P}_2 \boxtimes \mathbf{P}_1) \left[(\mathbf{A}_1 \odot \mathbf{B}_1) 1_{L_1}, \ldots, (\mathbf{A}_R \odot \mathbf{B}_R) 1_{L_R} \right]^T$ and $\mathbf{H}_2 = \left[(\mathbf{A}_1 \odot \mathbf{B}_1) 1_{L_1}, \ldots, (\mathbf{A}_R \odot \mathbf{B}_R) 1_{L_R} \right]^T$.

The updates for $\mathbf{P}_2$ (resp. $\mathbf{P}_3$) are obtained by substituting $\mathbf{P}_1$ by $\mathbf{P}_2$ (resp. $\mathbf{P}_3$), $\mathbf{Y}_1^{(1)}$ by $\mathbf{Y}_1^{(2)}$ (resp. $\mathbf{Y}_2^{(3)}$) and defining $\mathbf{V}^T = \mathbf{B}(\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A})^T$ (resp. $\mathbf{V}^T = \mathbf{C}(\mathbf{A} \odot_{\text{vec}} \mathbf{B})^T$) in (12).

Appendix B
Spatial degradation matrices

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

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

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

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

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

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

Appendix C
References

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

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

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

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

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

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

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

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

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

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

[11] Minghua Wang, Danfeng Hong, Zhu Han, Jiaxin Li, Jing Yao, Lianru Gao, Bing Zhang, and Jocelyn Chanussot, "Tensor decompositions for hyperspectral data processing in remote sensing: A comprehensive review," *arXiv preprint arXiv:2205.06407*, 2022.

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

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

[14] Clémence Prévost, Ricardo A. Borsoi, Konstantin Usevich, David Brie, José C. M. Bermudez, and Cédric Richard, "Hyperspectral super-resolution accounting for spectral variability: Coupled tensor ll1-based recovery and blind unmixing of the unknown super-resolution image," *SIAM Journal on Imaging Sciences*, vol. 15, no. 1, pp. 110–138, 2022.

[15] Hao Guo, Wenxing Bao, Kewen Qu, Xuan Ma, and Meng Cao, "Multispectral and hyperspectral image fusion based on regularized coupled non-negative block-term tensor decomposition," *Remote Sensing*, vol. 14, no. 21, pp. 5306, 2022.

[16] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y.Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, 2016.

[17] Valentin Leplat, Nicolas Gillis, and Andersen MS Ang, "Blind audio source separation with minimum-volume beta-divergence nmf," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.

[18] Sandro M Krieg, Niels H Buchmann, Jens Gempt, Ehab Shiban, Bernhard Meyer, and Florian Ringel, "Diffusion tensor imaging fiber tracking using navigated brain stimulation—a feasibility study," *Acta neurochirurgica*, vol. 154, no. 3, pp. 555–563, 2012.

[19] Charilaos I Kanatsoulis and Nicholas D Sidiropoulos, "Tex-graph: Coupled tensor-matrix knowledge-graph embedding for covid-19 drug repurposing," in *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*. SIAM, 2021, pp. 603–611.

[20] N. Gillis, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, 2020.

[21] V. Leplat, N. Gillis, and C. Févotte, "Multi-resolution beta-divergence nmf for blind spectral unmixing," *Signal Processing*, 2021.

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

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

[24] M. Bousse, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2016.

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

[26] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[27] Brian King, Cédric Févotte, and Paris Smaragdis, "Optimal cost function and magnitude power for nmf-based speech separation and music interpolation," in *2012 IEEE International Workshop on Machine Learning for Signal Processing*, 2012, pp. 1–6.

[28] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4810–4819, 2015.

[29] Y. Sun, P. Babu, and D.P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.

[30] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

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

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

[33] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1128–1141, 2015.

[34] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[35] Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.

[36] Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma, "Hyperspectral super-resolution via coupled tensor factorization: Identifiability and algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3191–3195.

[37] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Coupled tensor low-rank multilinear approximation for hyperspectral super-resolution," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 5536–5540.

[38] Ricardo A. Borsoi, Clémence Prévost, Konstantin Usevich, David Brie, José C. M. Bermudez, and Cédric Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 702–717, 2021.

[39] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

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

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