# DATA FUSION AND UNMIXING WITH THE REGULARIZED NON-NEGATIVE BLOCK-TERM DECOMPOSITION: JOINT PROBLEMS, BLIND APPROACH AND AUTOMATIC MODEL ORDER SELECTION

Clémence Prévost, Valentin Leplat

## ▶ To cite this version:

Clémence Prévost, Valentin Leplat. DATA FUSION AND UNMIXING WITH THE REGULARIZED NON-NEGATIVE BLOCK-TERM DECOMPOSITION: JOINT PROBLEMS, BLIND APPROACH AND AUTOMATIC MODEL ORDER SELECTION. 2023. hal-04302397

HAL Id: hal-04302397
https://hal.science/hal-04302397v1

Preprint submitted on 23 Nov 2023

# DATA FUSION AND UNMIXING WITH THE REGULARIZED NON-NEGATIVE BLOCK-TERM DECOMPOSITION: JOINT PROBLEMS, BLIND APPROACH AND AUTOMATIC MODEL ORDER SELECTION *

CLÉMENCE PRÉVOST† AND VALENTIN LEPLAT‡

**Abstract.** This paper introduces a family of coupled tensor optimization problems for joint super-resolution and unmixing in remote sensing. Using $\beta$-divergences allows the proposed methods to account for various noise statistics. A family of simple, efficient and flexible algorithms is proposed, that are capable of solving the two problems at hand. Moreover, the proposed algorithms are able to estimate the degradation operators mapping the HSI and MSI to the unknown SRI. We introduce penalized versions of our optimization problems, with an emphasis on the minimum-volume regularization approach. This approach is designed to efficiently identify the number of factors in the tensor decomposition, and to effectively manage scenarios involving potential rank deficiencies in the estimated mixing factors. It facilitates the computation of interpretable and meaningful tensor decompositions, and enhances the identifiability of the decomposition model. The proposed algorithms demonstrate competitive performance against state-of-the-art methods for joint fusion and unmixing, even in scenarios with various noise statistics and challenging cases, including partially unknown degradation operators, almost collinear materials, and estimation of the number of endmembers.

**Key words.** Nonnegative tensor factorization, block-term decomposition, $\beta$-divergence, minimum volume regularization, automatic model order selection, blind spectral unmixing, hyperspectral super-resolution.

**MSC codes.** 15A690, 90C26, 65F55

## 1. Introduction.

In remote sensing, hyperspectral images (HSIs) provide views of a portion of the Earth with high spectral resolution. However, due to the tradeoff between spatial and spectral resolutions, the HSIs to have low spatial resolution [39]. On the other hand, multispectral images (MSIs) possess high spatial resolution, at the cost of a restricted number of spectral bands. The composition of each pixel in these images can be approximated by a sum of a small number of spectral signatures, or endmembers. This representation is known as the linear mixing model. Blind spectral unmixing consists in identifying the materials present within the scenery with limited prior information, classically by computing the spectral signatures of these materials (the endmembers), and their abundance maps. However, due to the limited resolutions of the HSIs and MSIs, unmixing with high resolution cannot be performed on these images.

The hyperspectral super-resolution (HSR) problem [46] was formulated to circumvent the physical limitations of the HSIs and MSIs. It aims at recovering a super-resolution image (SRI) with both high spatial and high spectral resolutions from an HSI and an MSI of the same scene. Traditional unmixing can then be performed on the reconstructed SRI. Hence the goal of developing an efficient method for solving both problems at once.

Early matrix approaches for the HSR problem [40,44,45,47] performed a coupled low-rank factorization of the matricized HSI and MSI. Some of them were based on the linear mixing model, [27,47], thus they were suitable for joint HSR and unmixing.

More recently, tensor-based approaches were proposed for the HSR problem.

---

*Submitted to the editors November 6, 2023. This work is an extension of [33].

**Funding:** 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.

†Univ. Lille, UMR 9189 CRIStAL, F-59000 Lille, France (firstname.lastname[at]univ-lille.fr).;

‡Skoltech, Center for Artificial Intelligence Technology (CAIT), Moscow, Russia(V.Leplat[at]skoltech.ru).

In [17, 18], the HSR problem was formulated as a coupled canonical polyadic decomposition, while a coupled Tucker decomposition was used in [35]. However, the factors of these decompositions lacked physical interpretation, thus these methods could not perform unmixing.

The block-term decomposition (BTD) had been successfully used to perform unmixing on the SRI in [36]. Motivated by its usefulness, and by the success of tensor-based HSR, it has since then grown very popular to perform HSR [6,14,29,49]. In [32], the BTD was used for joint fusion and unmixing in the presence of spectral variability.

Successfully solving both the HSR and unmixing problems faces several challenges. First, the observation models and assorted algorithms must enforce non-negativity constraints. These constraints, when applied on the BTD factors, ensure that they can be recast in the linear mixing model as endmembers and abundance maps. The vast majority of algorithms for tensor/matrix non-negative factorization (NTF/NMF) rely on Block Coordinate Descent (BCD) schemes, see e.g., [4, 7, 15, 28].

Second, in remote sensing, observed images often contain highly similar or nearly collinear materials. For instance, consider the scenario where mineral components within a rock sample or materials like road, soil, and sand are primarily composed of silica. This property of the images may lead unmixing algorithms to recover rank-deficient mixing factors. This is particularly challenging when using the BTD, since its numerical implementation is very sensitive to the initialization, as highlighted in [32].

The third and most important challenge consists in correctly estimating the number of materials underlying the images. In practice, this value is often unknown. An incorrect estimation may negatively affect the performance of the fusion and unmixing algorithms, and may also not ensure the uniqueness of the solutions to these problems. In the absence of groundtruth reference, the number of materials can be estimated prior to processing owing to the extensive spectral information available in the HSI [1, 11]. Some recent works were devoted to the task of estimating the number of factors in BTD models using e.g., penalized optimization [38], Bayesian inference [12] or autoregressive models [48].

In this paper, a family of coupled optimization problems based on the BTD is introduced. The case of joint HSR and unmixing in remote sensing exemplifies the interest for these optimization problems. Other fields of applications could be envisioned for the considered model, such as audio signal processing [24], biomedical imaging [23] or graph signal processing [20]. Using $\beta$-divergences, various noise statistics present within the data can be accounted for, see [13] for a detailed overview of the topic. A family of simple, efficient and flexible algorithms is developed, that are able to handle the three above challenges within a unified framework.

Previously, the short paper [33] provided a brief overview of the optimization problem and partial simulations. This paper introduces a new family of regularized and constrained non-negative BTD problems and conduct additional simulations, including cases with collinear materials, unknown spatial degradation, and estimation of the model. Our main contributions are the following.

- We develop a family of tensor-based algorithms utilizing multiplicative updates tailored for the $\beta$-divergence. Among others, these algorithms allow the estimation of the degradation operators responsible for mapping the HSI and MSI to the unknown SRI. This fundamental principle, initially introduced in [25] through coupled matrix models, is extended into a coupled tensor framework.
- We incorporate a minimum-volume regularization into our optimization procedures. This regularization is designed to facilitate the computation of an interpretable and meaningful nonnegative BTD, while concurrently strengthening the identifiability of the decomposition model. It also addresses the challenge of rank deficiency and automating the critical step of determining

the number of factors, commonly referred to as “model order selection” in the matrix case. In this paper, we use this terminology when discussing the ability of a method to identify the factor count within the decomposition. It is imperative not to conflate this concept with the “order” of the input tensor, defined as its number of dimensions. While the advantages of the minimum-volume regularization were previously discussed in [24], this previous work exclusively applied this regularization to uncoupled matrix models.

- To circumvent the numerical sensitivity of the BTD, a robust way to initialize the algorithms based on matrix multiplicative updates is introduced.
- The proposed methods compete favorably with the state-of-the-art for joint fusion and unmixing of synthetic and semi-real data sets including three noise statistics: Gaussian noise, Poisson noise and multiplicative Gamma noise. Finally, we showcase the good performance of our methods in challenging cases: partially unknown degradation operators, almost collinear materials, and estimation of the number of endmembers.

This paper is organized as follows. Subsection 2.1 introduces the background on tensor algebra, the observation model and its assumptions. Section 3 describes the optimization problems that we consider. Section 4 contains the proposed algorithms and the suggested initialization strategy. Section 5 presents the numerical experiments on a series of synthetic and semi-real datasets.

## 2. Background and problem formulation.

**2.1. Background on tensor algebra.** The following notations [5,22] are used: lower ($a$) or uppercase ($A$) plain font for scalars, boldface lowercase ($\boldsymbol{a}$) 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 $\mathbf{A}^\mathsf{T}$. Notation $\mathbf{I}_N$ is used 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 $\mathbf{X} \geqslant \mathbf{0}$ 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 $\boxdot$. The operator 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.

Definition 2.1 introduces the BTD with ranks $(L_r, L_r, 1)$, that will be used to build the model. The main advantage of this decomposition is to link the low-rank factors 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_r$1-BTD) if*

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left(\mathbf{A}_r\mathbf{B}_r^\mathsf{T}\right) \otimes \boldsymbol{c}_r, \tag{2.1}$$

*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 $\boldsymbol{c}_r \in \mathbb{R}^K$, 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 $\mathbf{C} = [\boldsymbol{c}_1,\ldots,\boldsymbol{c}_R] \in \mathbb{R}^{K\times R}$.*

Property 2.2 recalls the unfolding formulae for the $L_rL_r$1-BTD, that will be helpful for building our algorithms.

Property 2.2 (Tensor unfoldings). *Using the above notation, the unfoldings of* *a tensor $\boldsymbol{\mathcal{X}}$ admitting an $L_rL_r1$-BTD as above can be expressed as*

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

*where $\odot_p$ denotes the partition-wise Khatri-Rao products defined as follows: $\mathbf{C}\odot_p\mathbf{A} = [\boldsymbol{c}_1\boxtimes\mathbf{A}_1,\dots,\boldsymbol{c}_R\boxtimes\mathbf{A}_R]$.*

**2.2. Assumptions.** Let us consider two tensors $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{I_1\times J_1\times K_1}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{I_2\times J_2\times K_2}$. The observed tensors $\boldsymbol{\mathcal{Y}}_1,\ \boldsymbol{\mathcal{Y}}_2$ are degraded versions of the same tensor $\boldsymbol{\mathcal{Y}}\in\mathbb{R}^{I\times J\times K}$. In remote sensing, the tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ respectively denote the HSI and MSI, and $\boldsymbol{\mathcal{Y}}$ denotes the unknown SRI we intend to recover. Thus we assume $I_1 < I_2,\ J_1 < J_2$ and $K_2 < K_1$: indices $I_\ell,\ J_\ell$ denote the spatial dimensions whereas $K_\ell$ denote the spectral ones ($\ell = 1,2$). In order to ease the notation, we assume that $I = I_2,\ J = J_2$ et $K = K_1$.

Assumption 2.3 (Structure of the SRI). *In the noiseless case, the tensor $\boldsymbol{\mathcal{Y}}$ admits a BTD with ranks $(L_r, L_r, 1)$ such that*

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R}\left(\mathbf{A}_r\mathbf{B}_r^{\mathsf{T}}\right)\otimes\boldsymbol{c}_r. \tag{2.2}$$

Under nonnegativity constraints, the vectors $\boldsymbol{c}_r$ in Equation (2.2) can be viewed as the endmembers associated to the $R$ constitutive materials of $\boldsymbol{\mathcal{Y}}$, while the matrices $\mathbf{A}_r\mathbf{B}_r^{\mathsf{T}} = \mathbf{S}_r \in \mathbb{R}^{I\times J}$ represent the corresponding abundance maps.

Assumption 2.4 (Structure of $\mathbf{S}_r$). *Matrices $\mathbf{S}_r$ are low-rank matrices,* i.e.*,*

$$\mathbf{S}_r \approx \mathbf{A}_r\mathbf{B}_r^{\mathsf{T}} \in \mathbb{R}^{I\times J}, \tag{2.3}$$

*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 $r\in\{1,\dots,R\}$.*

This assumption is reasonable, since an upper bound on the reconstruction error of such matrices by (2.3) can be obtained [3]. Furthermore, this assumptions will serve to link the linear mixing model to the $L_rL_r1$-BTD.

Let $\mathbf{S} = [\mathrm{vec}\{\mathbf{S}_1\},\dots,\mathrm{vec}\{\mathbf{S}_R\}] \in \mathbb{R}^{IJ\times R}$ be the matrix containing the vectorized abundance maps of each material and $\mathbf{C} = [\mathbf{c}_1,\dots,\mathbf{c}_R] \in \mathbb{R}^{L\times R}$ the matrix whose columns are the spectral signatures. The transposed third-mode unfolding of Equation (2.2) [6,49] reads

$$\mathbf{Y}^{(3)\mathsf{T}} = \mathbf{S}\mathbf{C}^{\mathsf{T}} \in \mathbb{R}^{IJ\times K}, \tag{2.4}$$

which is the linear mixing model (LMM) for the SRI $\boldsymbol{\mathcal{Y}}$ under nonnegativity constraints. Using Assumption 2.4 the block-term structure (2.2) can thus be viewed as tensor format for the LMM, under low-rank constraints of the abundance maps.

**2.3. Observation model.** As done in previous works (see [43] and references therein), the following model providing the links between $\boldsymbol{\mathcal{Y}}$ and its two degraded versions $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ is considered. The tensors $\boldsymbol{\mathcal{Y}}$ and ($\boldsymbol{\mathcal{Y}}_1$,$\boldsymbol{\mathcal{Y}}_2$) are such that

$$\begin{cases}\boldsymbol{\mathcal{Y}}_1 & \approx \sum\limits_{r=1}^{R}\left(\mathbf{P}_1\mathbf{A}_r\left(\mathbf{P}_2\mathbf{B}_r\right)^{\mathsf{T}}\right)\otimes\boldsymbol{c}_r,\\ \boldsymbol{\mathcal{Y}}_2 & \approx \sum\limits_{r=1}^{R}\left(\mathbf{A}_r\mathbf{B}_r^{\mathsf{T}}\right)\otimes\mathbf{P}_3\boldsymbol{c}_r,\end{cases} \tag{2.5}$$

which is a coupled $L_rL_r1$-BTD. The tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are degraded from $\boldsymbol{\mathcal{Y}}$ 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 [42]. The approximately equal symbols in Equation (2.5) account for the presence of noise during the degradation process.

## 3. Optimization problems.

### 3.1. Formulation of the joint fusion and unmixing problem.

State-of-the-art unmixing algorithms aim at recovering $\{\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^\top\}_{r=1}^R$ and $\mathbf{C}$ from the mixed pixels in $\boldsymbol{\mathcal{Y}}$. Here, since $\boldsymbol{\mathcal{Y}}$ is unknown and only $\boldsymbol{\mathcal{Y}}_1$ is observed with high spectral resolution, traditional algorithms are only able to recover spatially-degraded versions of the abundance maps [6], namely

$$\mathbf{P}_1\mathbf{S}_r\mathbf{P}_2^\top \in \mathbb{R}^{I_1 \times J_1} \text{ for } r \in \{1, \ldots, R\}. \tag{3.1}$$

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

Furthermore, in remote sensing, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ can be acquired at different time instants. The different acquisition conditions can result in e.g., variations in atmospheric, seasonal or illumination conditions [16]. Moreover, the specificities of the imaging devices might change after launch of installation in an aircraft, due to e.g., outgassing, aging of components, or misalignment. These uncertainties motivate the need for more flexible models, capable of estimating one or several degradation matrices $\mathbf{P}_i$ ($i \in \{1, \ldots, 3\}$).

Thus jointly solving the data fusion and blind unmixing problems consists in finding the ($L_rL_r1$) factors $\{\mathbf{A}_r\mathbf{B}_r^\top\}_{r=1}^R$, $\mathbf{C}$, and possibly the degradation matrices $\mathbf{P}_i$, under the assumption of (2.5), subject to the constraints

$$\{\mathbf{A}_r\mathbf{B}_r^\top\}_{r=1}^R \in \mathcal{X}_{\mathbf{A},\mathbf{B}},\ \mathbf{C} \in \mathcal{X}_{\mathbf{C}},\ \mathbf{P}_i \in \mathcal{X}_{\mathbf{P},i} \text{ for } i \in \{1, \ldots, 3\}, \tag{3.2}$$

where $\mathcal{X}_{\mathbf{T}} \subseteq \mathbb{R}^{M \times N}$ denotes the feasible (convex) set for matrix $\mathbf{T}$ of size $M \times N$.

### 3.2. Non-negative tensor optimization problem.

Trying to minimize the approximation errors in (2.5) leads, for instance, to minimizing the following general cost function:

$$\begin{aligned}\Phi =& D_\beta\left(\boldsymbol{\mathcal{Y}}_1 \| \sum_{r=1}^R \left(\mathbf{P}_1\mathbf{A}_r\left(\mathbf{P}_2\mathbf{B}_r\right)^\top\right) \otimes \boldsymbol{c}_r\right) + \lambda D_\beta\left(\boldsymbol{\mathcal{Y}}_2 \| \sum_{r=1}^R \left(\mathbf{A}_r\mathbf{B}_r^\top\right) \otimes \mathbf{P}_3\boldsymbol{c}_r\right) \\ &+ \sum_{r=1}^R \gamma_{(\mathbf{AB}),r}\Psi_{(\mathbf{AB}),r}(\mathbf{A}_r\mathbf{B}_r^\top) + \gamma_{\mathbf{C}}\Psi_{\mathbf{C}}(\mathbf{C}) + \sum_{i=1}^3 \gamma_{\mathbf{P}_i}\Psi_{\mathbf{P}_i}(\mathbf{P}_i)\end{aligned} \tag{3.3}$$

subject to the constraints in (3.2). The scalar $\lambda$ is a positive penalty parameter controlling the weights for the data fitting terms associated to each observation in the cost function $\Phi$, while $\gamma_{(.)}$ are positive parameters controlling the weights of the regularization functions $\Psi(.)$ potentially applied to all the factors of the $L_rL_r1$-BTD. Such regularization functions are used to promote solutions with desired and meaningful structures, such as low-rank, sparsity and minimum-volume to cite a few.

For a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$, the beta-divergence is defined as:

$$D_\beta\left(\boldsymbol{\mathcal{Y}} \| \sum_{r=1}^R \left(\mathbf{A}_r\mathbf{B}_r^\top\right) \otimes \boldsymbol{c}_r\right) = \sum_{i,j,k} d_\beta\left((\mathcal{Y})_{i,j,k} \| \left((\mathbf{A}_r)_{i,:}(\mathbf{B}_r)_{:,j}^\top\right) \otimes (\boldsymbol{c}_r)_k\right), \tag{3.4}$$

with $d_\beta(x\|y)$ the $\beta$-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 $\beta$-divergence corresponds to the Kullback-Leibler (KL) divergence and the Itakura-Saito (IS) divergence, respectively. For NMF and NTF models, the data fitting term should be chosen depending on the noise statistic assumed in the generative model of the data, see [7, 8, 21, 25] and references therein for more details. Section 4 presents our Algorithms to tackle specific instances of the general family of problems given in Equation (3.3).

**3.3. Considered variants of the problem.** In this paper, two specific instances of the optimization problem (3.3) subject to constraints in (3.2) are considered.

*Nonnegative and non-regularized.* For this case, $\gamma_{(.)} = 0$ for the regularization functions $\Psi(.)$, and the feasible sets $\mathcal{X}_{(.)}$ in (3.2) are the nonnegative orthants of appropriate dimensions. The optimization problem defined here will be dubbed as "$\beta$-$(L_r, L_r, 1)$-NBTD".

*Nonnegative minimum-volume.* The following family of optimization problems is considered:

(3.5)
$$\begin{aligned}\min_{\{\mathbf{S}_r\},\{\mathbf{P}_i\},\mathbf{C}} & D_\beta\left(\boldsymbol{\mathcal{Y}}_1 \| \sum_{r=1}^{R}\left(\mathbf{P}_1\mathbf{A}_r\left(\mathbf{P}_2\mathbf{B}_r\right)^\mathsf{T}\right)\otimes \boldsymbol{c}_r\right) + \lambda D_\beta\left(\boldsymbol{\mathcal{Y}}_2 \| \sum_{r=1}^{R}\left(\mathbf{A}_r\mathbf{B}_r^\mathsf{T}\right)\otimes \mathbf{P}_3\boldsymbol{c}_r\right) \\ & + \gamma \operatorname{vol}(\mathbf{C}) \\ & \text{s.t. } \mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^T \geqslant 0 \ \forall r, \mathbf{P}_i \geqslant \mathbf{0} \ \forall i, \boldsymbol{c}_r \in \Delta^K \ \forall r,\end{aligned}$$

where $\Delta^K = \{\boldsymbol{x} \in \mathbb{R}_+^K \,|\, \sum_{k=1}^K (\boldsymbol{x})_k = 1\}$, and $\operatorname{vol}(\mathbf{C})$ is a function measuring the volume spanned by the columns of $\mathbf{C}$. In [24], the authors use $\operatorname{vol}(\mathbf{C}) = \log\det(\mathbf{C}^T\mathbf{C} + \delta\mathbf{I})$ where $\delta$ is a small positive constant that prevents $\log\det(\mathbf{C}^T\mathbf{C})$ from going to $-\infty$ when $C$ tends to a rank deficient matrix (that is, when $\operatorname{rank}(\mathbf{C}) < R$). This model is particularly powerful as it leads to identifiability which is crucial in many applications such as in hyperspectral imaging or audio source separation [10]. Finally, the feasible sets for factors $\{\mathbf{A}_r\mathbf{B}_r^\mathsf{T}\}_{r=1}^R$ and $\mathbf{P}_i$ are the nonnegative orthant of appropriate size, while the feasible set for factor $\mathbf{C}$ is the set of column-stochastic matrices. The problem defined in (3.5) will be dubbed as "min-vol $\beta$-$(L_r, L_r, 1)$-NBTD".

Section 4 explains how to tackle the two variants of optimization problems detailed above. In particular, the optimization problem (3.5) is handled with the general framework presented in [26] with a special focus on the Kullback-Leibler (KL) divergence, that is when $\beta = 1$.

## 4. Algorithms.

This section addresses the $\beta$-$(L_r, L_r, 1)$-NBTD problem. Subsequently, we will proceed to present our algorithmic solution for tackling the "min-vol KL-$(L_r, L_r, 1)$-NBTD" problem outlined in (3.5).

**4.1. Nonnegative and non-regularized.** Most nonnegative tensor decomposition algorithms are based on an iterative scheme that alternatively updates one factor at the time with the others kept fixed, as it will be adopted in this paper. The goal in this section is to derive an algorithm to solve the nonnegative and non-regularized version of (3.3) based on the multiplicative updates (MU) rule. Let us consider the subproblem in $\mathbf{A}$ (with the others fixed) after unfolding along the first mode following Property 2.2:

$$\min_{A \geqslant \mathbf{0}} D_\beta(\mathbf{Y}_1^{(1)} \| \mathbf{P}_1\mathbf{A}(\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})^\mathsf{T}) + \lambda D_\beta(\mathbf{Y}_2^{(1)} \| \mathbf{A}(\mathbf{P}_3\mathbf{C} \odot_p \mathbf{B})^\mathsf{T}). \tag{4.1}$$

To tackle this problem, we follow the standard majorization-minimization (MM) framework [41] and the results given by [25, Lemma 2]. Given the current iterate

$\tilde{\mathbf{A}}$, let us pose $\mathbf{H}_1 = (\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})^\mathsf{T}$ and $\mathbf{H}_2 = (\mathbf{P}_3\mathbf{C} \odot_p \mathbf{B})^\mathsf{T}$, we obtain the following update:

$$\mathbf{A} = \tilde{\mathbf{A}} \boxdot \left( \frac{\left[\mathbf{P}_1^T \left( \left(\mathbf{P}_1\tilde{\mathbf{A}}\mathbf{H}_1\right)^{.(\beta-2)} \boxdot \mathbf{Y}_1^{(1)} \right) \mathbf{H}_1^T + \lambda \left( \left(\tilde{\mathbf{A}}\mathbf{H}_2\right)^{.(\beta-2)} \boxdot \mathbf{Y}_2^{(1)} \right) \mathbf{H}_2^T \right]}{\left[\mathbf{P}_1^T \left(\mathbf{P}_1\tilde{\mathbf{A}}\mathbf{H}_1\right)^{.(\beta-1)} \mathbf{H}_1^T + \lambda \left(\tilde{\mathbf{A}}\mathbf{H}_2\right)^{.(\beta-1)} \mathbf{H}_2^T \right]} \right)^{.\gamma(\beta)} \tag{4.2}$$

where $A \boxdot B$ (resp. $\frac{[A]}{[B]}$) is the Hadamard product (resp. division) between $A$ and $B$, $A^{(.\alpha)}$ is the element-wise $\alpha$ 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 $\beta > 2$ [9]. The subproblems in $\mathbf{B}$ and $\mathbf{C}$ can be solved similarly and their closed form expressions can be found in Appendix A.

As opposed to the majority of state-of-the-art methods, the proposed 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 $\mathbf{X} = \mathbf{U}\mathbf{V}^\mathsf{T}$ [9]. For $\mathbf{P}_1$, we are interested in solving $D_\beta(\mathbf{Y}_1^{(1)} \| \mathbf{P}_1\mathbf{A}(\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})^\mathsf{T})$. By posing $\mathbf{V}^\mathsf{T} = \mathbf{A}(\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})^\mathsf{T}$, we derive:

$$\mathbf{P}_1 \leftarrow \widetilde{\mathbf{P}}_1 \boxdot \left( \frac{\left[ \left( \widetilde{\mathbf{P}}_1\mathbf{V}^\mathsf{T} \right)^{.(\beta-2)} \boxdot \mathbf{Y}_1^{(1)} \right] \mathbf{V}}{\left[ \widetilde{\mathbf{P}}_1\mathbf{V}^\mathsf{T} \right]^{.(\beta-1)} \mathbf{V}} \right)^{.\gamma(\beta)}. \tag{4.3}$$

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

Algorithm 4.1 summarizes the proposed method to tackle the $\beta$-($L_r$,$L_r$,1)-NBTD optimization problem (3.3). 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 [25] 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 $\Phi$ from (3.3) is below some given threshold $\kappa$, or when the maximum number of iterations is reached.

**Initialization:** Many options are available to initialize factors $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. In this paper, an efficient way to initialize the low-rank factors of the $L_rL_r1$-BTD is proposed. For factor $\mathbf{C}$, VCA [31] is performed 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 under nonnegative constraints:

$$\min_{\mathbf{S} \geqslant 0} \ \mathbf{P}_3\mathbf{C}\mathbf{S}^\mathsf{T} \approx \mathbf{Y}_2^{(3)}. \tag{4.4}$$

Initialization of $\mathbf{A}_r$ and $\mathbf{B}_r$ for $r \in \{1, \dots, R\}$ is performed based on the classical MU $\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^\mathsf{T}$ [9] with a maximum of `j1` iterations.

In the semi-blind case, *i.e.*, when one or several matrices $\mathbf{P}_i$ are unknown, they are initialized similarly using (4.3) with a maximum of `j2` iterations. The initialization procedure is summarized in Algorithm 4.2.

**Algorithm 4.1** MU for $\beta$-($L_r$,$L_r$,1)-NBTD

```
Require: Input tensors $\boldsymbol{\mathcal{Y}}_1 \geqslant \mathbf{0}$, $\boldsymbol{\mathcal{Y}}_2 \geqslant \mathbf{0}$, initializations $\mathbf{A}, \mathbf{B}, \mathbf{C} \geqslant \mathbf{0}$, downsampling operators $\mathbf{P}_i \geqslant \mathbf{0}$ 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 $\lambda > 0$.
Ensure: An approximate solution to (3.3) under constraints (3.2)
% Loop 1: update of matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ only
$i \leftarrow 0$, $\Phi^0 = 1$, $\Phi^1 = 0$.
while $i <$ i1 and $\left|\frac{\Phi^i - \Phi^{i+1}}{\Phi^i}\right| > \kappa$ do
    Update $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ sequentially; see Equations (4.2)
    Compute the objective function $\Phi^{i+1}$
end while
% Loop 2 : update of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{P}_i$
$i \leftarrow 0$
while $i <$ i2 and $\left|\frac{\Phi^i - \Phi^{i+1}}{\Phi^i}\right| > \kappa$ do
    Update $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{P}_i$ for $i \in \{1, ..., 3\}$ sequentially; see Equations (4.2) and (4.3)
    Compute the objective function $\Phi^{i+1}$
end while
return $\hat{\boldsymbol{\mathcal{Y}}} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}\right) \otimes \boldsymbol{c}_r$
```

**Algorithm 4.2** Initialization of Algorithm 4.1

```
Require: Input tensors $\boldsymbol{\mathcal{Y}}_1 \geqslant \mathbf{0}$, $\boldsymbol{\mathcal{Y}}_2 \geqslant \mathbf{0}$, downsampling operators $\mathbf{P}_i \geqslant \mathbf{0}$ for $i \in \{1, ..., 3\}$; $R$, ranks $\{L_r\}_{r=1}^{R}$, maximum number of iterations j1 and j2, a threshold $0 < \kappa \ll 1$.
Ensure: Initial values $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$
Initialize $\mathbf{C}$ using VCA on $\boldsymbol{\mathcal{Y}}_1$;
Compute $\mathbf{S}$ using result of (4.4);
% 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 $\mathbf{A}_r$, $\mathbf{B}_r$ using the classical MU
    Compute the objective function $\Phi^{j+1}$
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 $\mathbf{A}_r$, $\mathbf{B}_r$, and $\mathbf{P}_i$ using matrix-based MU
    Compute the objective function $\Phi^{j+1}$
end while
return $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_R]$, $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_R]$ and $\mathbf{C}$
```

**4.2. Nonnegative minimum-volume.** The primary approach for addressing the "min-vol KL-($L_r, L_r$, 1)-NBTD" problem in equation (3.5) is analogous to the methodology employed in Section 4.1, specifically the cyclic block majorization minimization (BMM) framework. This framework involves cyclically updating each block of variables with the others kept fixed, with each update accomplished by minimizing a majorizer constructed using the current iterate. The MU for updating the factors

$\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^{R}$ and $\{\mathbf{P}_i\}_{i=1}^{3}$ for Problem (3.5) are the same as the ones presented in the previous section dedicated to tackling the $\beta$-($L_r$,$L_r$,1)-NBTD problem, since the sub-problems in theses factors are identical. Therefore, this subsection focuses on solving the sub-problem of (3.5) that concerns the minimum-volume penalty and column-stochasticity constraints, specifically the updating factor $\mathbf{C}$. The recent optimization framework proposed in [26] is used to derive efficient MU for the sub-problem in $\mathbf{C}$, which is defined as follows:

$$\begin{aligned} &\min_{\mathbf{C}} D_\beta(\mathbf{Y}_1^{(3)} \| \mathbf{C}\mathbf{H}_1) + \lambda D_\beta(\mathbf{Y}_2^{(3)} \| \mathbf{P}_3\mathbf{C}\mathbf{H}_2) + \gamma \log\det(\mathbf{C}^T\mathbf{C} + \delta\mathbf{I}) \\ &\qquad \text{s.t. } e^T\mathbf{C} = e^T, \mathbf{C} \geqslant \mathbf{0}. \end{aligned} \tag{4.5}$$

with $\mathbf{H}_1 = (\mathbf{P}_1\mathbf{A} \odot_{\text{vec}} \mathbf{P}_2\mathbf{B})^{\mathsf{T}}$, and $\mathbf{H}_2 = (\mathbf{A} \odot_{\text{vec}} \mathbf{B})^{\mathsf{T}}$.

An upper bound for the term $\log\det(\mathbf{C}^T\mathbf{C}+\delta\mathbf{I})$ is required by [26, Assumption. 1]. It is majorized using a convex quadratic separable auxiliary function provided in [24, Eq. (3.6)] and which is derived as follows. First, the concave function $\log\det(\mathbf{Q})$ for $\mathbf{Q} \succ \mathbf{0}$ can be upper bounded using the first-order Taylor approximation: for any $\widetilde{\mathbf{Q}} \succ \mathbf{0}$,

$$\log\det(\mathbf{Q}) \;\leqslant\; \log\det(\widetilde{\mathbf{Q}}) + \langle \widetilde{\mathbf{Q}}^{-1}, \mathbf{Q} - \widetilde{\mathbf{Q}} \rangle \;=\; \langle \widetilde{\mathbf{Q}}^{-1}, \mathbf{Q} \rangle + \text{cst},$$

where cst is some constant independent of $\mathbf{Q}$. For any $\mathbf{C}, \widetilde{\mathbf{C}}$, and denoting $\widetilde{\mathbf{Q}} = \widetilde{\mathbf{C}}^T\widetilde{\mathbf{C}} + \delta\mathbf{I} \succ \mathbf{0}$, we obtain

$$\log\det(\mathbf{C}^T\mathbf{C} + \delta\mathbf{I}) \;\leqslant\; \left\langle \widetilde{\mathbf{Q}}^{-1}, \mathbf{C}^T\mathbf{C} \right\rangle + \text{cst} \;=\; \text{trace}(\mathbf{C}\widetilde{\mathbf{Q}}^{-1}\mathbf{C}^T) + \text{cst}, \tag{4.6}$$

which is a convex quadratic and Lipschitz-smooth function in $\mathbf{C}$, and where $\widetilde{\mathbf{C}}$ will be practically chosen as the current iterate for $\mathbf{C}$. With this and following the framework from [26], the Lagrangian function is built as follows

$$\begin{aligned} G^{\mu}(\mathbf{C}|\widetilde{\mathbf{C}}) = &\sum_{k}^{K} G_{Y_1}(\boldsymbol{c}_k|\widetilde{\boldsymbol{c}}_k) + \lambda \sum_{k}^{K} G_{Y_2}(\boldsymbol{c}_k|\widetilde{\boldsymbol{c}}_k) + \gamma \left( \sum_{k}^{K} \bar{l}(\boldsymbol{c}_k|\widetilde{\boldsymbol{c}}_k) + \text{c} \right) \\ &+ \boldsymbol{\mu}^T \sum_{k}^{K} \left( \boldsymbol{c}_k - \frac{1}{K}\boldsymbol{e} \right), \end{aligned} \tag{4.7}$$

where $\boldsymbol{c}_k$ denotes the $k$-th row of $\mathbf{C}$, $G$ is the separable majorizer for $\beta$-divergence proposed in [9] (one for each $\beta$-divergence $D_\beta$ in (4.5)), $\bar{l}$ is a convex and separable majorizer[1] for the convex quadratic (4.6) and given by [24, Eq. (3.6)], and c is a constant. Let $\boldsymbol{\mu}$ be the vector of Lagrange multipliers of dimension $R$ associated to each linear constraint $\boldsymbol{e}^T\boldsymbol{c}_r = 1$. One can easily show that $G^{\mu}$ is separable w.r.t. each component of $\boldsymbol{c}_k$ and, given $\boldsymbol{\mu}$, one can compute the closed-form solution (the details are omitted):

$$\mathbf{C}^{\star}(\mu) = \widetilde{\mathbf{C}} \odot \frac{\left[ \left[ \left[ \mathbf{Q} + \boldsymbol{e}\boldsymbol{\mu}^T \right]^{.2} + \mathbf{R} \right]^{.\frac{1}{2}} - (\mathbf{Q} + \boldsymbol{e}\boldsymbol{\mu}^T) \right]}{[\mathbf{D}]}, \tag{4.8}$$

where $\mathbf{Q} = \mathbf{E}_{K,I_1J_1}\mathbf{S}^T - 4\gamma(\widetilde{\mathbf{C}}\mathbf{U}^-) + \lambda\mathbf{P}_3^T\mathbf{E}_{K_2,IJ}\mathbf{H}^T$, $\mathbf{D} = 4\gamma\widetilde{\mathbf{C}}(\mathbf{U}^+ + \mathbf{U}^-)$, and $\mathbf{R} = 8\gamma\widetilde{\mathbf{C}}(\mathbf{U}^+ + \mathbf{U}^-) \odot \left( \frac{[\mathbf{Y}_1^{(3)}]}{[\widetilde{\mathbf{C}}\mathbf{H}_1]}\mathbf{H}_1^T + \lambda\mathbf{P}_3^T \frac{[\mathbf{Y}_2^{(3)}]}{[\mathbf{P}_3\widetilde{\mathbf{C}}\mathbf{H}_2]}\mathbf{H}_2^T \right)$ with $\mathbf{U} = \mathbf{U}^+ - \mathbf{U}^- =$

---

[1] tight at the current iterate $\widetilde{\mathbf{C}}$, that is $\sum_k^K \bar{l}(\widetilde{\boldsymbol{c}}_k|\widetilde{\boldsymbol{c}}_k) = \text{trace}(\widetilde{\mathbf{C}}\widetilde{\mathbf{Q}}^{-1}\widetilde{\mathbf{C}}^T)$

$\left(\widetilde{\mathbf{C}}^T\widetilde{\mathbf{C}}+\delta\mathbf{I}\right)^{-1}$, $\mathbf{U}^+=\max(\mathbf{U},\mathbf{0})\geqslant\mathbf{0}$ and $\mathbf{U}^-=\max(-\mathbf{U},\mathbf{0})\geqslant\mathbf{0}$, and $\mathbf{E}_{K,I_1J_1}$ is the $K$-by-$I_1J_1$ matrix of all ones. As proved in [26, Proposition. 2], the constraint $e^T\mathbf{C}^\star(\mu)=e^T$ is satisfied for a unique $\boldsymbol{\mu}\in\mathbb{R}^R$. Therefore a Newton-Raphson method can be used to find the $\boldsymbol{\mu}$ with quadratic rate of convergence; see [26, Proposition. 3]. Algorithm 4.3 summarizes our method to tackle (3.5).

**Algorithm 4.3** MU for min-vol KL-($L_r$,$L_r$,1)-NBTD

**Require:** Input tensors $\boldsymbol{\mathcal{Y}}_1\geqslant\mathbf{0}$, $\boldsymbol{\mathcal{Y}}_2\geqslant\mathbf{0}$, initializations $\mathbf{A},\mathbf{B},\mathbf{C}\geqslant\mathbf{0}$, downsampling operators $\mathbf{P}_i\geqslant\mathbf{0}$ 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 weights $\lambda,\gamma>0$.

**Ensure:** An approximate solution to (3.5)

```
% 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 A, B
    Update A, B and C sequentially; see Equations (4.2)
    % Update of matrix C
    $\mathbf{U} \leftarrow \left(\mathbf{C}^T\mathbf{C}+\delta\mathbf{I}\right)^{-1}$
    $\mathbf{U}^+ \leftarrow \max(\mathbf{U},0)$
    $\mathbf{U}^- \leftarrow \max(-\mathbf{U},0)$
    $\mathbf{Q} \leftarrow \mathbf{E}_{K,I_1J_1}\mathbf{H}_1^T - 4\gamma\left(\mathbf{C}\mathbf{U}^-\right) + \lambda\mathbf{P}_3^T\mathbf{E}_{K_2,IJ}\mathbf{H}_2^T$
    $\mathbf{D} \leftarrow 4\gamma\mathbf{C}\left(\mathbf{U}^+ + \mathbf{U}^-\right)$
    $\mathbf{R} \leftarrow 8\gamma\mathbf{C}\left(\mathbf{U}^+ + \mathbf{U}^-\right) \odot \left(\frac{\left[\mathbf{Y}_1^{(3)}\right]}{[\mathbf{C}\mathbf{H}_1]}\mathbf{H}_1^T + \lambda\mathbf{P}_3^T\frac{\left[\mathbf{Y}_2^{(3)}\right]}{[\mathbf{P}_3\mathbf{C}\mathbf{H}_2]}\mathbf{H}_2^T\right)$
    $\boldsymbol{\mu} \leftarrow \mathrm{root}\left(e^T\mathbf{C}^\star(\mu) = e^T\right)$ over $\mathbb{R}^R$    % see (4.8) for the expression of $\mathbf{C}^\star(\mu)$
    $\mathbf{C} \leftarrow \mathbf{C} \odot \frac{\left[\left[\left[\mathbf{Q}+e\boldsymbol{\mu}^T\right]^{.2}+\mathbf{R}\right]^{.\frac{1}{2}}-\left(\mathbf{Q}+e\boldsymbol{\mu}^T\right)\right]}{[\mathbf{D}]}$
    Compute the objective function $\Phi^{i+1}$
end while
% Loop 2
$i \leftarrow 0$
while $i <$ i2 and $\left|\frac{\Phi^i-\Phi^{i+1}}{\Phi^i}\right| > \kappa$ do
    Update A, B, C and $\mathbf{P}_i$ for $i \in \{1,...,3\}$ sequentially; see Equations (4.2),
        (4.3) and update of C as performed in Loop 1
    Compute the objective function $\Phi^{i+1}$
end while
return $\hat{\boldsymbol{\mathcal{Y}}} = \sum_{r=1}^R \left(\mathbf{A}_r\mathbf{B}_r^\mathsf{T}\right) \otimes \boldsymbol{c}_r$
```

### 4.3. Key insights about the Algorithms.

The following paragraphs discuss key elements for Algorithms 4.1 and 4.3 proposed in previous sections.

**Comments on computational complexity:** The computational complexity of Algorithm 4.1, designed to solve the $\beta$-($L_r$, $L_r$, 1)-NBTD problem, exhibits asymptotic equivalence to the standard MU for $\beta$-NMF, after unfoldings. This implies that it requires $\mathcal{O}(IJK\times\sum_r^R L_r)$ operations per iteration. The main driver of this complexity lies in the aforementioned matrix products. One might question the influence of the computation step of Algorithm 4.3, which aims to calculate the optimal Lagrangian multipliers, thereby ensuring the column-stochasticity constraints of matrix $\mathbf{C}$ are satisfied. In [26], the authors demonstrate the quadratic convergence of the procedure employed to compute the Lagrangian multipliers. As a result, the number of iterations required for this process remains low in practical scenarios. Consequently, the overall complexity of Algorithm 4.3 is predominantly determined by the computationally intensive matrix products mentioned earlier. For more comprehensive information, we refer the reader to [26].

**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 the MU given in Equation (4.2), 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, the updates of factors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{P}_i$ are obtained from 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 [9, 25], and a small positive scalar $\epsilon$ (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 [37].

## 5. Numerical experiments.

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 online at https://github.com/cprevost4/bLL1_NBTD.

### 5.1. Test setup.

The proposed algorithms were compared to several tensor methods designed for solving the HSR problem, namely STEREO and Blind-STEREO [17, 19], SCOTT and BSCOTT [34], CT-STAR and CB-STAR [2], SCLL1 [6] and CNN-BTD-Var [32]. Among them, SCLL1 and CNN-BTD-Var were based on the LL1-BTD model[2], therefore they are able to solve the unmixing problem[3].

Several matrix-based approaches were also benchmarked: CNMF [47], FUSE [46], HySure [40], SFIM [46] and MR-$\beta$-NMF [25]. Being based on coupled nonnegative matrix factorization, CNMF and MR-$\beta$-NMF were able to perform joint fusion and unmixing. In particular, MR-$\beta$-NMF was based on multiplicative updates with the $\beta$-divergence, but did not include a volume-regularizing constraint. The ranks and regularization parameters were chosen according to the original works.

The proposed algorithms were initialized using Algorithm 4.2 with a maximum of 500 iterations and a threshold $\kappa = 10^{-7}$. All methods were limited to a maximum of 1000 iterations, encompassing both loops, and utilized a fixed value of $\kappa = 10^{-7}$.

The groundtruth SRI $\boldsymbol{\mathcal{Y}}$ was compared to the estimated SRI $\widehat{\boldsymbol{\mathcal{Y}}}$ obtained by the algorithms. The main performance metric used in comparisons was the *Peak Signal-to-Noise ratio* (PSNR) [46]:

$$\text{PSNR} = 10\log_{10}\left(\frac{\|\boldsymbol{\mathcal{Y}}\|_F^2}{\|\widehat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}\|_F^2}\right). \tag{5.1}$$

In addition to PSNR, we considered different metrics [46] described below:

$$\text{CC} \ = \frac{1}{IJK}\left(\sum_{k=1}^{K} \rho\left(\boldsymbol{\mathcal{Y}}_{:,:,k}, \widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}\right)\right), \tag{5.2}$$

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

---

[2] These works all considered the simpler model with $L_1 = \ldots = L_R$.

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

spectral slices;

$$\text{ERGAS} \;=\frac{100}{d}\sqrt{\frac{1}{IJK}\sum_{k=1}^{K}\frac{\|\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}-\boldsymbol{\mathcal{Y}}_{:,:,k}\|_F^2}{\mu_k^2}}, \tag{5.3}$$

where $\mu_k^2$ is the mean value of $\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}$. 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):

$$\text{SAD} \;=\frac{1}{R}\sum_{r=1}^{R}\arccos\left(\frac{\mathbf{c}_r^{\mathsf{T}}\widehat{\mathbf{c}}_r}{\|\mathbf{c}_r\|_2\|\widehat{\mathbf{c}}_r\|_2}\right), \tag{5.4}$$

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 $\widehat{\mathbf{S}}$:

$$\text{RMSE} \;=\frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{IJ}\sum_{d=1}^{IJ}\Big((\mathbf{S}_r)_d-(\widehat{\mathbf{S}}_r)_d)\Big)^2}. \tag{5.5}$$

The computational time for each algorithm was given by the `tic` and `toc` functions of Matlab. Finally, the compression rate (C.R.) of each method was computed, *i.e.*, the ratio between the number of elements in the unknown SRI and the number of unknown parameters in the algorithms. The higher this value, the less parameters need to be recovered by the model.

**5.2. Degradation model.** We considered synthetic and semi-real datasets for which a reference SRI is available. The HSI was obtained by spatial degradation of $\boldsymbol{\mathcal{Y}}$ using $\mathbf{P}_1$ and $\mathbf{P}_2$ while the MSI was obtained by spectral degradation of $\boldsymbol{\mathcal{Y}}$ with $\mathbf{P}_3$ according to model (2.5).

For spatial degradation, we followed the commonly used Wald's protocol [42]. The matrices $\mathbf{P}_1$, $\mathbf{P}_2$ were computed with a separable Gaussian blurring kernel of size $q = 9$. Downsampling was performed along each spatial dimension with a ratio $d = 4$ between the SRI and HSI, as in previous works [17]– [6].

For the spectral degradation matrix $\mathbf{P}_3$, the spectral response functions of the Sentinel-2 instrument[4] was used. It spans the electromagnetic spectrum from 412nm to 2022nm and produces a 10-band MSI ($K_2 = 10$) 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 $\mathbf{P}_3$ acts as a selection-weighting matrix, utilizing the shared spectral bands between the SRI and the MSI datasets.

As done in [25], three noise statistics were considered: Gaussian Noise, Poisson noise and Multiplicative Gamma noise. Gaussian and Poisson noise were added to the observations in order to yield 30dB SNR. For Gamma noise, we considered a distribution of mean 1 and variance 0.05. Since both observations are subject to the same level of noise, we fixed the weight $\lambda = 1$.

---

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

**5.3. Datasets.** Experiments were run on synthetic and semi-real datasets described below.

**Synthetic dataset:** A dataset where the SRI admits an exact $L_rL_r1$-BTD was created, in order to perform simulations in a controlled environment.

We considered $R = 4$ spectral signatures $\mathbf{c}_r$ ($r \in \{1, \ldots, R\}$) obtained from the Jasper Ridge reference data[5]. The SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ ($I = J = 120$, $K = 173$) 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 $\frac{I}{L_r} \times \frac{J}{L_r}$. Each block in the parcel map was a patch composed of entries equal to one. At most one material is active in each block, thus the so-called "pure pixel assumption" was valid. Figure 1 depicts the abundance maps for the four materials in the synthetic data set.

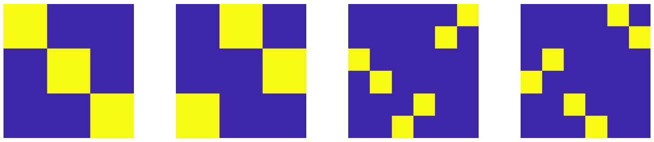


FIG. 1. *Abundance maps for the four materials in the synthetic dataset.*

The reference SRI is computed as follows:

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

**Semi-real datasets:** The first semi-real dataset we considered is built upon the Jasper Ridge reference SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{100\times100\times173}$. It includes four materials: road, soil, water and vegetation. We chose $R = 4$, and $L_1 = 15$, $L_2 = 8$, $L_3 = 20$, $L_4 = 13$.

A second semi-real dataset was considered and based on the Ivanpah Playa [30] reference SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{95\times95\times156}$ This dataset is composed of four, very similar materials: road, solar panels, light sand and darker sand. Thus it constitutes a challenging example for which unmixing may lead to rank-deficient solutions.

**5.4. Reconstruction results.** In order to assess the performance of our approaches for data fusion, the quality metrics obtained with each method over 5 trials are reported in the Tables 1–3 below. The two best metrics of each columns are shown in bold. For comparison with Algorithm 4.3, we only show the results for the KL divergence, *i.e.*, for $\beta = 1$ along with Poisson noise. The results for other noise statistics are available in supplementary materials.

The proposed algorithms yield better metrics than most baseline methods. Moreover, they provide excellent reconstruction results for, e.g., the Ivanpah Playa dataset. The compression rate is higher than that of matrix-based approaches (*i.e.*, less information was needed to reconstruct the SRI). It is however lower than that of other tensor-based algorithms, but provide significantly better results than, e.g., CT-STAR and CNN-BTD-Var for the semi-real datasets.

As it will be highlighted in the following sections, the good unmixing performance of the minimim-volume regularized approach come at the cost of a tradeoff in the reconstruction performance. The proposed algorithms however remain very competitive for reconstruction of the three considered SRI.

**5.5. Unmixing results.** For the unmixing task, the reference and estimated spectral signatures and abundance maps are depicted. Due to space limitations,

---

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

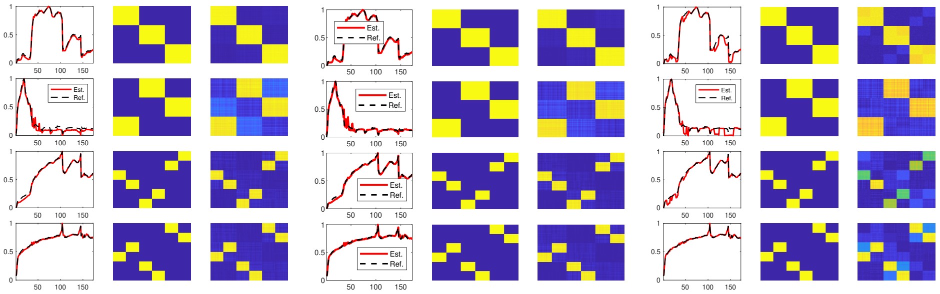
| Method | CC (↑) | SAD (↓) | RMSE (↓) | ERGAS (↓) | PSNR (dB) (↑) | Time (sec) (↓) | C.R. (↑) |
|---|---|---|---|---|---|---|---|
| STEREO | 0.999 | 2.82 | 8.882e-3 | 1.347 | 35.31 | 1.573 | 120 |
| BSTEREO | 0.995 | 7.98 | 2.134e-2 | 2.809 | 29.93 | 1.379 | 101 |
| SCOTT | 0.999 | 2.77 | 1.122e-2 | 1.518 | 34.30 | 2.087 | 234 |
| BSCOTT | 0.999 | 1.47 | 1.146e-2 | 1.601 | 34.11 | 0.204 | 159 |
| SCLL1 | 0.987 | 5.70 | 3.609e-2 | 5.434 | 24.46 | 33.57 | **266** |
| CT-STAR | **0.999** | **0.25** | **5.272e-3** | **0.713** | **39.71** | **0.102** | 242 |
| CB-STAR | **0.999** | 1.25 | **6.408e-3** | 0.991 | **38.02** | 20.14 | 138 |
| CNN-BTD-Var | 0.997 | 5.30 | 1.260e-2 | 2.952 | 32.76 | 3.302 | 266 |
| CNMF | 0.999 | **1.09** | 7.313e-3 | 1.364 | 36.20 | 1.872 | 43 |
| FUSE | 0.994 | 2.77 | 1.470e-2 | 2.813 | 30.92 | **0.146** | 6 |
| HySure | 0.998 | 3.04 | 1.375e-2 | 2.174 | 32.35 | 17.48 | 43 |
| SFIM | 0.990 | 4.96 | 2.544e-2 | 3.817 | 28.07 | 0.281 | 6 |
| MR-$\beta$(=1)-NMF | 0.998 | 2.22 | 7.026e-2 | 1.581 | 37.08 | 586.77 | 43 |
| Alg. 4.1 ($\beta = 1$) | 0.999 | 1.19 | 1.065e-2 | **0.776** | 35.69 | 23.42 | **497** |
| Alg. 4.3 ($\beta = 1$) | 0.999 | 2.96 | 8.703e-3 | 2.007 | 36.88 | 30.19 | **497** |

TABLE 1
*Reconstruction of the synthetic dataset with Poisson noise.*

| Method | CC (↑) | SAD (↓) | RMSE (↓) | ERGAS (↓) | PSNR (dB) (↑) | Time (sec) (↓) | C.R. (↑) |
|---|---|---|---|---|---|---|---|
| STEREO | 0.989 | 3.01 | 1.378e-2 | 2.617 | **31.10** | 1.597 | 92 |
| BSTEREO | 0.989 | 3.05 | 1.435e-2 | 2.479 | 30.69 | 1.673 | 82 |
| SCOTT | 0.991 | 3.30 | 1.584e-2 | 2.361 | 30.20 | **0.180** | 191 |
| BSCOTT | 0.981 | 3.55 | 2.282e-2 | 3.191 | 28.11 | 0.213 | 127 |
| SCLL1 | 0.979 | 5.23 | 2.983e-2 | 3.656 | 25.46 | 11.387 | 199 |
| CT-STAR | 0.886 | 8.63 | 5.328e-2 | 7.204 | 19.53 | **0.092** | **456** |
| CB-STAR | 0.992 | 2.89 | 1.371e-2 | 2.269 | 31.09 | 7.592 | 116 |
| CNN-BTD-Var | 0.920 | 8.63 | 4.152e-2 | 6.078 | 21.45 | 1.130 | **199** |
| CNMF | **0.999** | **0.90** | **6.000e-3** | **1.493** | 27.96 | 1.302 | 42 |
| FUSE | 0.984 | 2.88 | 2.066e-2 | 3.227 | 28.15 | 0.276 | 8 |
| HySure | **0.998** | **2.05** | **1.196e-2** | **1.778** | **33.55** | 13.734 | 42 |
| SFIM | 0.989 | 3.71 | 2.083e-2 | 2.832 | 29.54 | 0.320 | 8 |
| MR-$\beta$(=1)-NMF | 0.964 | 5.63 | 2.111e-2 | 6.514 | 27.81 | 61.268 | 42 |
| Alg. 4.1 ($\beta = 1$) | 0.990 | 3.16 | 1.712e-2 | 2.657 | 29.58 | 42.241 | 87 |
| Alg. 4.3 ($\beta = 1$) | 0.986 | 3.95 | 2.144e-2 | 2.88 | 27.58 | 19.77 | 87 |

TABLE 2
*Reconstruction metrics on the Jasper Ridge dataset, Poisson noise.*

only the results for our algorithms and SCLL1 in the scenario of Poisson additive noise are shown. The other results, including the other baseline algorithms and noise statistics, are available as supplementary materials. Figures 2–4 show the reference and estimated spectra (first column), reference abundance maps (second column) and the estimated maps (third column).



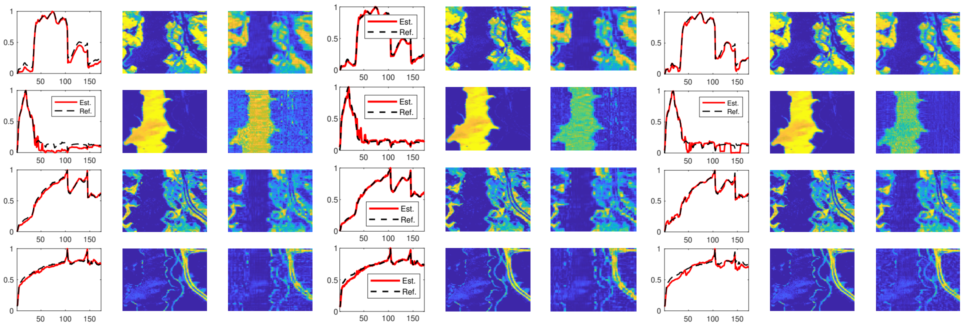
(a) (b) (c)

FIG. 2. *Unmixing on the synthetic dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.*

| Method | CC (↑) | SAD (↓) | RMSE (↓) | ERGAS (↓) | PSNR (dB) (↑) | Time (sec) (↓) | C.R. (↑) |
|---|---|---|---|---|---|---|---|
| STEREO | 0.987 | 1.82 | 1.421e-2 | 1.119 | 32.07 | 1.640 | 92 |
| BSTEREO | 0.979 | 2.84 | 1.897e-2 | 1.477 | 29.90 | 1.499 | 81 |
| SCOTT | 0.979 | 1.80 | 1.629e-2 | 1.265 | 30.58 | 0.189 | **196** |
| BSCOTT | 0.870 | 5.45 | 3.946e-2 | 3.320 | 23.82 | **0.115** | 142 |
| SCLL1 | 0.974 | 5.12 | 6.996e-2 | 3.985 | 22.05 | 2.838 | 86 |
| CT-STAR | 0.948 | 1.30 | 2.257e-2 | 1.595 | 27.01 | **0.053** | **442** |
| CB-STAR | 0.988 | 1.24 | 1.215e-2 | 0.969 | 32.31 | 5.955 | 196 |
| CNN-BTD-Var | 0.973 | **1.11** | 1.827e-2 | 1.295 | 28.42 | 2.727 | 86 |
| CNMF | **0.988** | 1.12 | **8.598e-3** | **0.766** | **34.96** | 1.931 | 42 |
| FUSE | **0.991** | **1.05** | **9.621e-3** | **0.825** | **34.02** | 0.233 | 8 |
| HySure | 0.939 | 3.71 | 2.583e-2 | 2.083 | 27.15 | 13.852 | 42 |
| SFIM | 0.963 | 2.08 | 1.605e-2 | 1.289 | 30.02 | 0.258 | 8 |
| MR-$\beta$(=1)-NMF | 0.983 | 1.82 | 1.415e-2 | 1.573 | 32.26 | 152.81 | 42 |
| Alg. 4.1 ($\beta = 1$) | 0.983 | 1.75 | 1.489e-2 | 1.176 | 31.43 | 46.472 | 125 |
| Alg. 4.3 ($\beta = 1$) | 0.984 | 1.74 | 1.502e-2 | 1.172 | 31.33 | 19.16 | 125 |

TABLE 3

*Reconstruction metrics on the Ivanpah Playa dataset, Poisson noise.*



(a) (b) (c)

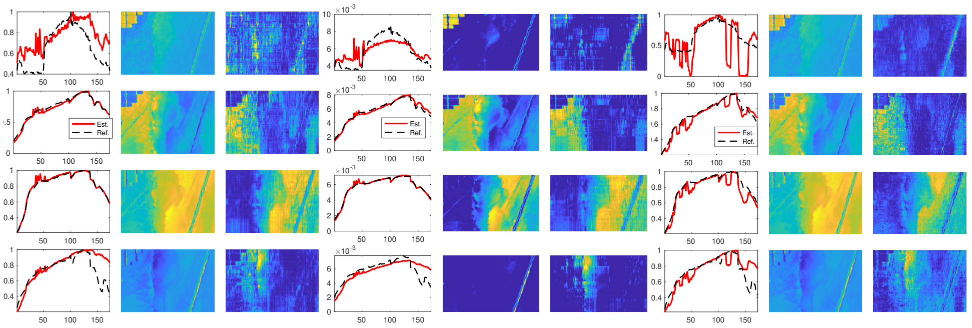
FIG. 3. *Unmixing on the Jasper Ridge dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.*

One can observe that the baseline algorithm SCLL1 provide rather poor results, due to its incapacity to account for different noise statistics and different $L_r$ values. Conversely, the spectral signatures and abundance maps are correctly estimated by the proposed algorithms for the three considered datasets. Algorithm 4.3 generally yields better estimation of the spectra and less artifacts on the abundance maps. Even though Ivanpah Playa is a notoriously difficult dataset for unmixing, Algorithm 4.3 provides a good estimation of the materials in the scene.

**5.6. Performance in a semi-blind scenario.** The performance of our algorithms was then evaluated in a semi-blind scenario, *i.e.*, when the spatial degradation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are unknown. This scenario is likely to occur, e.g., when the HSI and MSI are acquired at different times or in the presence of motion blur. The proposed algorithms were compared to two semi-blind tensor algorithms: BSTEREO and BSCOTT, that do not estimate $\mathbf{P}_1$ and $\mathbf{P}_2$.

We considered the Jasper Ridge dataset and the three noise statistics. Table 4 presents the reconstruction metrics obtained for the benchmarked algorithms. The proposed algorithms generally provides good reconstruction, especially with additive Gaussian noise. One can observe that using Algorithm 4.3 with $\beta = 1$ improves the overall reconstruction performance. Although the computation time is higher than that of the other tensor algorithms, it is significantly smaller than the one of the MR-$\beta$-NMF approach.

Figure 5 shows a portion of the reference and estimated spatial degradation (that



(a) (b) (c)

Fig. 4. *Unmixing on the Ivanpah Playa dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.*

| Method | CC (↑) | SAD (↓) | RMSE (↓) | ERGAS (↓) | PSNR (dB) (↑) | Time (sec) (↓) |
|---|---|---|---|---|---|---|
| **Gaussian Noise - 30 dB** | | | | | | |
| BSTEREO | 0.962 | 5.28 | 2.585e-2 | 4.690 | 25.13 | 0.974 |
| BSCOTT | 0.947 | 5.60 | 3.169e-2 | 5.972 | 24.48 | **0.108** |
| MR-$\beta$-NMF | 0.971 | 6.88 | 4.508e-2 | 6.120 | 22.34 | 1393.747 |
| Alg. 4.1 ($\beta = 2$) | **0.976** | **3.85** | **2.521e-2** | **4.407** | **26.36** | 74.142 |
| **Poisson Noise - 30dB** | | | | | | |
| BSTEREO | 0.962 | 5.67 | 2.726e-2 | 3.917 | 24.90 | 1.752 |
| BSCOTT | 0.951 | 5.57 | 3.094e-2 | 4.467 | 24.46 | 0.465 |
| MR-$\beta$-NMF | **0.988** | 4.57 | 2.306e-2 | 3.228 | **28.83** | 1187.941 |
| Alg. 4.1 ($\beta = 1$) | 0.964 | 6.95 | 2.653e-2 | 3.922 | 25.92 | 50.776 |
| Alg. 4.3 ($\beta = 1$) | 0.986 | **4.34** | **2.161e-2** | **2.937** | 27.76 | 26.59 |
| **Gamma Noise** | | | | | | |
| BSTEREO | 0.961 | 5.57 | **2.691e-2** | 4.503 | 24.82 | 1.689 |
| BSCOTT | 0.951 | 5.80 | 3.258e-2 | 4.936 | 24.13 | **0.483** |
| MR-$\beta$-NMF | **0.981** | 5.41 | 3.335e-2 | **4.46** | 25.58 | 1401.120 |
| Alg. 4.1 ($\beta = 0$) | 0.967 | **3.66** | 3.024e-2 | 5.627 | **25.93** | 90.296 |

Table 4
*Semi-blind reconstruction on the Jasper Ridge dataset.*

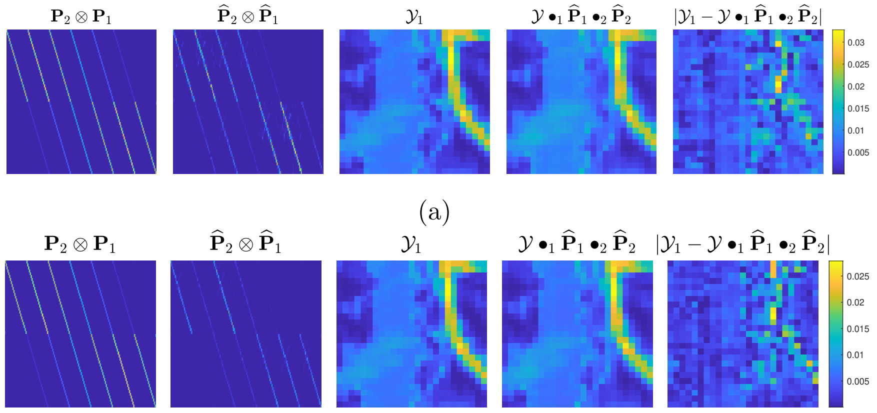
is, $\mathbf{P}_2 \boxtimes \mathbf{P}_1$). A single slice of the true HSI and the tensor constructed by mode product of the SRI with the estimated spatial degradation is also shown. The proposed algorithms recovers correctly the structure of the spatial degradation operators. Moreover, the HSI obtained from the estimated degradation is coherent with the reference HSI, and Algorithm 4.3 provides slightly better results.

Figure 6 shows our unmixing results in the case $\beta = 1$. The spectra and abundance maps are correctly estimated by the proposed algorithms, even without prior knowledge on the spatial degradation matrix. Resorting to Algorithm 4.3 offers better estimation of the spectral signatures and less artifacts due to the low-rank assumption on the abundance maps.

**5.7. Estimation of the number of endmembers.** This subsection highlights the capabilities of Algorithm 4.3 to accurately retrieve the number of endmembers underlying an image.

This test considered the synthetic dataset with 4 endmembers. We deliberately chose $R = 6$ for our algorithms to simulate an overestimation of the number of materials. The first 4 columns of $\mathbf{C}$ were initialized using VCA while the remaining two were initialized randomly. We chose $L_1 = L_2 = L_5 = L_6 = 3$ and $L_3 = L_4 = 6$.

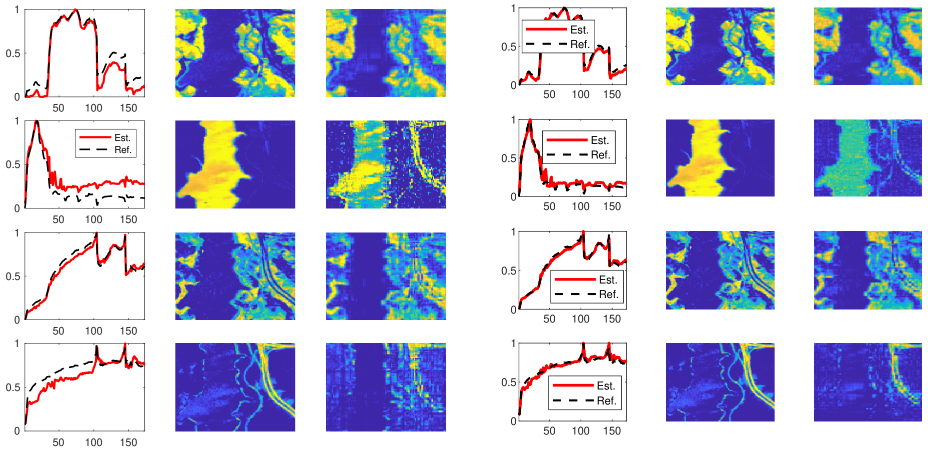
Figure 7 presents the unmixing results obtained with Algorithm 4.1 and Algo-



(a)

(b)

FIG. 5. *Estimation of the spatial degradation and comparison with the real HSI, (a) Algorithm 4.1 ($\beta = 1$), (b) Algorithm 4.3 ($\beta = 1$).*

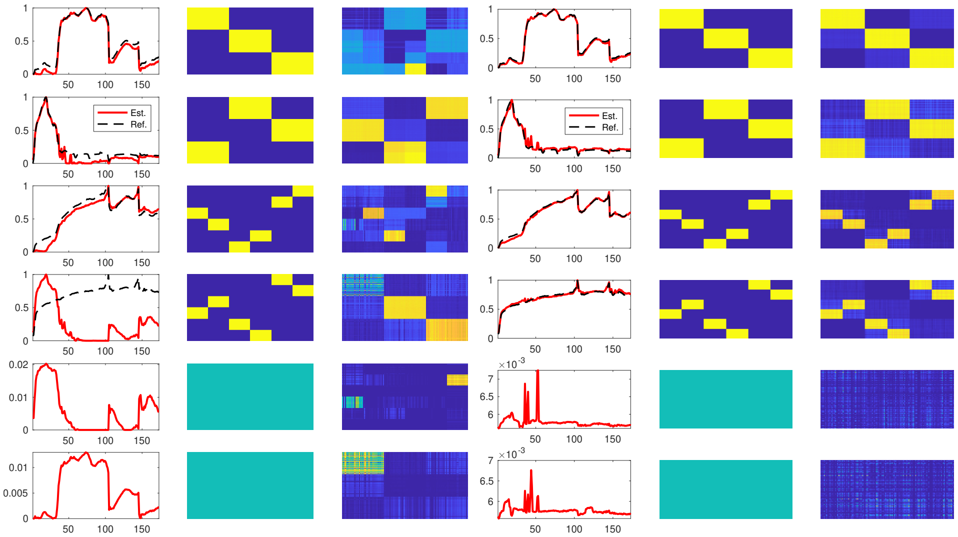


(a) (b)

FIG. 6. *Semi-blind unmixing, (a) Algorithm 4.1 ($\beta = 1$), (b) Algorithm 4.3 ($\beta = 1$).*

rithm 4.3 with $\beta = 1$. Algorithm 4.1 only recovers the first three spectra correctly, while the other three do not match the reference spectral signatures. The estimated abundance maps seem like mixtures of several reference abundance maps, thus they are not estimated correctly. Algorithm 4.3 recovers with high accuracy the four spectral signatures and abundance maps of interest. The other two materials are almost constant and show very low magnitude. These numerical evidences support the capacity of Algorithm 4.3 to estimate the correct number of materials, hence performing automatic model order selection, on top of the reconstruction and unmixing tasks.

**6. Conclusion.** This paper proposes a family of coupled tensor-based optimization problems. Using the beta-divergence, various noise statistics can be accounted for. The use of the BTD allows the proposed algorithms to jointly solve the reconstruction and unmixing tasks in remote sensing, even in challenging cases such as a semi-blind scenario or almost collinear materials. We introduced a new family of penalized optimization problems with a focus on the minimum-volume regularization. A



(a) (b)

FIG. 7. *Unmixing results with overestimation of the rank $R$, (a) Algorithm 4.1 ($\beta = 1$), (b) Algorithm 4.3 ($\beta = 1$).*

new Algorithm, dubbed as "min-vol $\beta$-($L_r$,$L_r$,1)-NBTD", was introduced to efficiently solve the regularized problems. The experiments show that the minimum-volume regularization approach improves the unmixing performance, at the cost of a small loss in the reconstruction performance. More importantly, the "min-vol $\beta$-($L_r$,$L_r$,1)-NBTD" algorithm correctly estimates the number of materials in the datasets, additionally to perform high-quality unmixing. The minimum-volume regularization approach also previously showed its interest for automatic model order selection in audio signal processing. Future works will be devoted to the mathematical analysis of this mechanism.

**Appendix A. Detailed MU updates.** This appendix contains the closed-form expressions for the updates of $\mathbf{B}$ and $\mathbf{C}$, similarly to (4.2).

(A.1)

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

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

(A.2)

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

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

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 $\mathbf{V}^{\mathsf{T}} = \mathbf{B}(\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A})^{\mathsf{T}}$ (resp. $\mathbf{V}^{\mathsf{T}} = \mathbf{C}(\mathbf{A} \odot_{\text{vec}} \mathbf{B})^{\mathsf{T}}$) in (4.3).

## REFERENCES

[1] J. M. Bioucas-Dias and J. P. Nascimento, *Hyperspectral subspace identification*, IEEE Trans. Geosci. Remote Sens., 46 (2008), pp. 2435–2445.

[2] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. M. Bermudez, and C. Richard, *Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability*, IEEE Journal of Selected Topics in Signal Processing, 15 (2021), pp. 702–717, https://doi.org/10.1109/JSTSP.2021.3054338.

[3] M. Bousse, O. Debals, and L. De Lathauwer, *A tensor-based method for large-scale blind source separation using segmentation*, IEEE Trans. Signal Process., 65 (2016), pp. 346–358.

[4] R. Bro and N. D. Sidiropoulos, *Least squares algorithms under unimodality and non-negativity constraints*, Journal of Chemometrics: A Journal of the Chemometrics Society, 12 (1998), pp. 223–247.

[5] P. Comon, *Tensors: A brief introduction*, IEEE Signal Process. Mag., 31 (2014), pp. 44–53.

[6] 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, (2020), arXiv:2006.10248, p. arXiv:2006.10248, https://arxiv.org/abs/2006.10248.

[7] C. Févotte, N. Bertin, and J.-L. Durrieu, *Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis*, Neural computation, 21 (2009), pp. 793–830.

[8] C. Févotte and N. Dobigeon, *Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization*, IEEE Transactions on Image Processing, 24 (2015), pp. 4810–4819.

[9] C. Févotte and J. Idier, *Algorithms for nonnegative matrix factorization with the $\beta$-divergence*, Neural computation, 23 (2011), pp. 2421–2456.

[10] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, *Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications.*, IEEE Signal Process. Mag., 36 (2019), pp. 59–80.

[11] 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., 9 (2015), pp. 1128–1141.

[12] P. V. Giampouras, A. A. Rontogiannis, and E. Kofidis, *Block-term tensor decomposition model selection and computation: The bayesian way*, IEEE Transactions on Signal Processing, 70 (2022), pp. 1704–1717.

[13] N. Gillis, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, 2020, https://doi.org/10.1137/1.9781611976410.

[14] H. Guo, W. Bao, K. Qu, X. Ma, and M. Cao, *Multispectral and hyperspectral image fusion based on regularized coupled non-negative block-term tensor decomposition*, Remote Sensing, 14 (2022), p. 5306.

[15] K. Huang, N. D. Sidiropoulos, and A. P. Liavas, *A flexible and efficient algorithmic framework for constrained matrix and tensor factorization*, IEEE Trans. Signal Process., 64 (2016), pp. 5052–5065.

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

[17] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, *Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach*, IEEE Trans. Signal Process., 66 (2018), pp. 6503–6517.

[18] 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.

[19] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. 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.

[20] C. I. Kanatsoulis and N. 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.

[21] B. King, C. Févotte, and P. 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, https://doi.org/10.1109/MLSP.2012.6349726.

[22] T. G. Kolda and B. W. Bader, *Tensor Decompositions and Applications*, SIAM Review, 51 (2009), pp. 455–500.

[23] S. M. Krieg, N. H. Buchmann, J. Gempt, E. Shiban, B. Meyer, and F. Ringel, *Diffusion tensor imaging fiber tracking using navigated brain stimulation—a feasibility study*, Acta neurochirurgica, 154 (2012), pp. 555–563.

[24] V. Leplat, N. Gillis, and A. M. Ang, *Blind audio source separation with minimum-volume beta-divergence nmf*, IEEE Transactions on Signal Processing, 68 (2020), pp. 3400–3410.

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

[26] V. Leplat, N. Gillis, and J. Idier, *Multiplicative updates for nmf with $\beta$-divergences under*

*disjoint equality constraints*, SIAM Journal on Matrix Analysis and Applications, 42 (2021), pp. 730–752.

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

[28] C.-J. Lin, *Projected gradient methods for nonnegative matrix factorization*, Neural computation, 19 (2007), pp. 2756–2779.

[29] H. Liu, W. Jiang, Y. Zha, and Z. Wei, *Coupled tensor block term decomposition with superpixel-based graph laplacian regularization for hyperspectral super-resolution*, Remote Sensing, 14 (2022), p. 4520.

[30] S. Mahalingam, P. Srinivas, P. K. Devi, D. Sita, S. K. Das, T. S. Leela, and V. R. Venkataraman, *Reflectance based vicarious calibration of hysis sensors and spectral stability study over pseudo-invariant sites*, in 2019 IEEE Recent Advances in Geoscience and Remote Sensing: Technologies, Standards and Applications (TENGARSS), IEEE, 2019, pp. 132–136.

[31] J. Nascimento and J. Dias, *Vertex component analysis: A fast algorithm to unmix hyperspectral data*, IEEE Geosci. Remote Sens. Lett., 43 (2005), pp. 898–910.

[32] C. Prévost, R. A. Borsoi, K. Usevich, D. Brie, J. M. Bermudez, and C. 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, 15 (2022), pp. 110–138, https://doi.org/10.1137/21M1409354.

[33] C. Prévost and V. Leplat, *Nonnegative block-term decomposition with the $\beta$-divergence: Joint data fusion and blind spectral unmixing*, in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

[34] 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.

[35] 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., 68 (2020), pp. 931–946.

[36] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Tang, *Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery*, IEEE Trans. Geosci. Remote Sens., 55 (2016), pp. 1776–1792.

[37] M. Razaviyayn, M. Hong, and Z.-Q. Luo, *A unified convergence analysis of block successive minimization methods for nonsmooth optimization*, SIAM Journal on Optimization, 23 (2013), pp. 1126–1153.

[38] A. A. Rontogiannis, E. Kofidis, and P. V. Giampouras, *Online rank-revealing block-term tensor decomposition*, Signal Processing, (2023), p. 109126.

[39] G. A. Shaw and H. K. Burke, *Spectral imaging for remote sensing*, Lincoln laboratory journal, 14 (2003), pp. 3–28.

[40] 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., 53 (2015), pp. 3373–3388.

[41] Y. Sun, P. Babu, and D. Palomar, *Majorization-minimization algorithms in signal processing, communications, and machine learning*, IEEE Transactions on Signal Processing, 65 (2017), pp. 794–816.

[42] 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., 63 (1997), pp. 691–699.

[43] M. Wang, D. Hong, Z. Han, J. Li, J. Yao, L. G., B. Zhang, and J. Chanussot, *Tensor decompositions for hyperspectral data processing in remote sensing: A comprehensive review*, arXiv preprint arXiv:2205.06407, (2022).

[44] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, *Multiband image fusion based on spectral unmixing*, IEEE Trans. Geosci. Remote Sens., 54 (2016), pp. 7236–7249.

[45] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, *Fast fusion of multi-band images based on solving a Sylvester equation*, IEEE Trans. Image Process., 24 (2015), pp. 4109–4121.

[46] N. Yokoya, C. Grohnfeldt, and J. Chanussot, *Hyperspectral and multispectral data fusion: A comparative review of the recent literature*, IEEE Trans. Geosci. Remote Sens., 5 (2017), pp. 29–56.

[47] N. Yokoya, T. Yairi, and A. Iwasaki, *Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion*, IEEE Trans. Geosci. Remote Sens., 50 (2012), pp. 528–537, https://doi.org/10.1109/TGRS.2011.2161320.

[48] V. Zarzoso, *Parameter estimation in block term decomposition for noninvasive atrial fibrillation analysis*, in 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2017, pp. 1–5.

[49] 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.