



**HAL**  
open science

# 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-04302397>**

Preprint submitted on 23 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

5                    CLÉMENTINE PRÉVOST<sup>†</sup> AND VALENTIN LEPLAT<sup>‡</sup>

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

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

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

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

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

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

44        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.

<sup>†</sup>Univ. Lille, UMR 9189 CRISTAL, F-59000 Lille, France (firstname.lastname[at]univ-lille.fr).;

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

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

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

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

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

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

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

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

- 87 • We develop a family of tensor-based algorithms utilizing multiplicative up-  
 88 dates tailored for the  $\beta$ -divergence. Among others, these algorithms allow the  
 89 estimation of the degradation operators responsible for mapping the HSI and  
 90 MSI to the unknown SRI. This fundamental principle, initially introduced  
 91 in [25] through coupled matrix models, is extended into a coupled tensor  
 92 framework.
- 93 • We incorporate a minimum-volume regularization into our optimization pro-  
 94 cedures. This regularization is designed to facilitate the computation of an  
 95 interpretable and meaningful nonnegative BTD, while concurrently strength-  
 96 ening the identifiability of the decomposition model. It also addresses the  
 97 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 BTM, 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 ( $\mathbf{a}$ ) for vectors, boldface uppercase ( $\mathbf{A}$ ) for matrices and calligraphic ( $\mathcal{A}$ ) for tensors. The elements of vectors, matrices and tensors are denoted as  $a_i$ ,  $A_{i,j}$  and  $\mathcal{A}_{i_1, \dots, i_N}$ , respectively. The transpose of a matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}^\top$ . 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  $\mathbf{1}_L$  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  $\square$ . The operator  $\text{vec}$  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 BTM 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  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$  admits a block-term decomposition (BTM) with ranks  $(L_r, L_r, 1)$  ( $L_r L_r 1$ -BTM) if*

$$(2.1) \quad \mathcal{X} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{c}_r,$$

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  $r \in \{1, \dots, R\}$ . Moreover, we denote  $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_R] \in \mathbb{R}^{I \times \sum_r L_r}$ ,  $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_R] \in \mathbb{R}^{J \times \sum_r L_r}$  and  $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$ .

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

PROPERTY 2.2 (Tensor unfoldings). *Using the above notation, the unfoldings of*

145 a tensor  $\mathcal{X}$  admitting an  $L_r L_r 1$ -BTD as above can be expressed as

$$\begin{aligned} 146 \quad \mathbf{X}^{(1)} &= \mathbf{A} (\mathbf{C} \odot_p \mathbf{B})^\top, \quad \mathbf{X}^{(2)} = \mathbf{B} (\mathbf{C} \odot_p \mathbf{A})^\top, \\ 147 \quad \mathbf{X}^{(3)} &= \mathbf{C} [(\mathbf{A}_1 \odot \mathbf{B}_1) \mathbf{1}_{L_1}, \dots, (\mathbf{A}_R \odot \mathbf{B}_R) \mathbf{1}_{L_R}]^\top, \end{aligned}$$

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

151 **2.2. Assumptions.** Let us consider two tensors  $\mathcal{Y}_1 \in \mathbb{R}^{I_1 \times J_1 \times K_1}$  and  $\mathcal{Y}_2 \in$   
152  $\mathbb{R}^{I_2 \times J_2 \times K_2}$ . The observed tensors  $\mathcal{Y}_1, \mathcal{Y}_2$  are degraded versions of the same tensor  
153  $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ . In remote sensing, the tensors  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$  respectively denote the HSI  
154 and MSI, and  $\mathcal{Y}$  denotes the unknown SRI we intend to recover. Thus we assume  
155  $I_1 < I_2, J_1 < J_2$  and  $K_2 < K_1$ : indices  $I_\ell, J_\ell$  denote the spatial dimensions whereas  
156  $K_\ell$  denote the spectral ones ( $\ell = 1, 2$ ). In order to ease the notation, we assume that  
157  $I = I_2, J = J_2$  et  $K = K_1$ .

158 ASSUMPTION 2.3 (Structure of the SRI). *In the noiseless case, the tensor  $\mathcal{Y}$  ad-*  
159 *mits a BTD with ranks  $(L_r, L_r, 1)$  such that*

$$160 \quad (2.2) \quad \mathcal{Y} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{c}_r.$$

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

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

$$166 \quad (2.3) \quad \mathbf{S}_r \approx \mathbf{A}_r \mathbf{B}_r^\top \in \mathbb{R}^{I \times J},$$

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

169 This assumption is reasonable, since an upper bound on the reconstruction error of  
170 such matrices by (2.3) can be obtained [3]. Furthermore, this assumptions will serve  
171 to link the linear mixing model to the  $L_r L_r 1$ -BTD.

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

$$176 \quad (2.4) \quad \mathbf{Y}^{(3)\top} = \mathbf{S} \mathbf{C}^\top \in \mathbb{R}^{I \times J \times K},$$

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

181 **2.3. Observation model.** As done in previous works (see [43] and references  
182 therein), the following model providing the links between  $\mathcal{Y}$  and its two degraded  
183 versions  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$  is considered. The tensors  $\mathcal{Y}$  and  $(\mathcal{Y}_1, \mathcal{Y}_2)$  are such that

$$184 \quad (2.5) \quad \begin{cases} \mathcal{Y}_1 & \approx \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^\top) \otimes \mathbf{c}_r, \\ \mathcal{Y}_2 & \approx \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{P}_3 \mathbf{c}_r, \end{cases}$$

185

186 which is a coupled  $L_r L_r 1$ -BTD. The tensors  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$  are degraded from  $\mathcal{Y}$  using  
 187 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  
 188 be full-rank. In a remote sensing framework, the matrix  $\mathbf{P}_3 \in \mathbb{R}^{K_2 \times K}$  contains the  
 189 spectral response functions for each band of the MSI sensor. The spatial degradation  
 190 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  
 191 along each spatial dimension, *i.e.* we suppose that the spatial degradation operation  
 192 is separable, as in the commonly used Wald's protocol [42]. The approximately equal  
 193 symbols in Equation (2.5) account for the presence of noise during the degradation  
 194 process.

### 195 3. Optimization problems.

196 **3.1. Formulation of the joint fusion and unmixing problem.** State-of-the-  
 197 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  
 198 pixels in  $\mathcal{Y}$ . Here, since  $\mathcal{Y}$  is unknown and only  $\mathcal{Y}_1$  is observed with high spectral  
 199 resolution, traditional algorithms are only able to recover spatially-degraded versions  
 200 of the abundance maps [6], namely

$$201 \quad (3.1) \quad \mathbf{P}_1 \mathbf{S}_r \mathbf{P}_2^T \in \mathbb{R}^{I_1 \times J_1} \text{ for } r \in \{1, \dots, R\}.$$

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

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

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

$$215 \quad (3.2) \quad \{\mathbf{A}_r \mathbf{B}_r^T\}_{r=1}^R \in \mathcal{X}_{\mathbf{A}, \mathbf{B}}, \quad \mathbf{C} \in \mathcal{X}_{\mathbf{C}}, \quad \mathbf{P}_i \in \mathcal{X}_{\mathbf{P}, i} \text{ for } i \in \{1, \dots, 3\},$$

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

218 **3.2. Non-negative tensor optimization problem.** Trying to minimize the  
 219 approximation errors in (2.5) leads, for instance, to minimizing the following general  
 220 cost function:

$$221 \quad (3.3) \quad \Phi = D_\beta \left( \mathcal{Y}_1 \parallel \sum_{r=1}^R \left( \mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r^T)^T \right) \otimes \mathbf{c}_r \right) + \lambda D_\beta \left( \mathcal{Y}_2 \parallel \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{P}_3 \mathbf{c}_r \right) \\ + \sum_{r=1}^R \gamma_{(\mathbf{A}, \mathbf{B}), r} \Psi_{(\mathbf{A}, \mathbf{B}), r}(\mathbf{A}_r \mathbf{B}_r^T) + \gamma_{\mathbf{C}} \Psi_{\mathbf{C}}(\mathbf{C}) + \sum_{i=1}^3 \gamma_{\mathbf{P}_i} \Psi_{\mathbf{P}_i}(\mathbf{P}_i)$$

222 subject to the constraints in (3.2). The scalar  $\lambda$  is a positive penalty parameter  
 223 controlling the weights for the data fitting terms associated to each observation in  
 224 the cost function  $\Phi$ , while  $\gamma_{(\cdot)}$  are positive parameters controlling the weights of  
 225 the regularization functions  $\Psi(\cdot)$  potentially applied to all the factors of the  $L_r L_r 1$ -  
 226 BTD. Such regularization functions are used to promote solutions with desired and  
 227 meaningful structures, such as low-rank, sparsity and minimum-volume to cite a few.

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

$$229 \quad (3.4) \quad D_\beta \left( \mathcal{Y} \parallel \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{c}_r \right) = \sum_{i,j,k} d_\beta ((\mathcal{Y})_{i,j,k} \parallel ((\mathbf{A}_r)_{i,:} (\mathbf{B}_r^T)_{:,j}) \otimes (\mathbf{c}_r)_k),$$

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,  $\gamma_{(\cdot)} = 0$  for the regularization functions  $\Psi(\cdot)$ , and the feasible sets  $\mathcal{X}_{(\cdot)}$  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:

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

where  $\Delta^K = \{\mathbf{x} \in \mathbb{R}_+^K \mid \sum_{k=1}^K (\mathbf{x})_k = 1\}$ , and  $\text{vol}(\mathbf{C})$  is a function measuring the volume spanned by the columns of  $\mathbf{C}$ . In [24], the authors use  $\text{vol}(\mathbf{C}) = \log \det(\mathbf{C}^\top \mathbf{C} + \delta \mathbf{I})$  where  $\delta$  is a small positive constant that prevents  $\log \det(\mathbf{C}^\top \mathbf{C})$  from going to  $-\infty$  when  $\mathbf{C}$  tends to a rank deficient matrix (that is, when  $\text{rank}(\mathbf{C}) < R$ ). 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  $\{\mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R$  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:

$$(4.1) \quad \min_{\mathbf{A} \geq 0} D_\beta(\mathbf{Y}_1^{(1)} \|\mathbf{P}_1 \mathbf{A} (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B})^\top) + \lambda D_\beta(\mathbf{Y}_2^{(1)} \|\mathbf{A} (\mathbf{P}_3 \mathbf{C} \odot_p \mathbf{B})^\top).$$

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

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

(4.2)

$$276 \quad \mathbf{A} = \tilde{\mathbf{A}} \square \left( \frac{\left[ \mathbf{P}_1^T \left( (\mathbf{P}_1 \tilde{\mathbf{A}} \mathbf{H}_1)^{(\beta-2)} \square \mathbf{Y}_1^{(1)} \right) \mathbf{H}_1^T + \lambda \left( (\tilde{\mathbf{A}} \mathbf{H}_2)^{(\beta-2)} \square \mathbf{Y}_2^{(1)} \right) \mathbf{H}_2^T \right]^{\cdot \gamma(\beta)}}{\left[ \mathbf{P}_1^T (\mathbf{P}_1 \tilde{\mathbf{A}} \mathbf{H}_1)^{(\beta-1)} \mathbf{H}_1^T + \lambda (\tilde{\mathbf{A}} \mathbf{H}_2)^{(\beta-1)} \mathbf{H}_2^T \right]^{\cdot \gamma(\beta)}} \right)$$

277 where  $A \square B$  (resp.  $\frac{A}{B}$ ) is the Hadamard product (resp. division) between  $A$  and  
 278  $B$ ,  $A^{(\cdot \alpha)}$  is the element-wise  $\alpha$  exponent of  $A$ ,  $\gamma(\beta) = \frac{1}{2-\beta}$  for  $\beta < 1$ ,  $\gamma(\beta) = 1$  for  
 279  $\beta \in [1, 2]$  and  $\gamma(\beta) = \frac{1}{\beta-1}$  for  $\beta > 2$  [9]. The subproblems in  $\mathbf{B}$  and  $\mathbf{C}$  can be solved  
 280 similarly and their closed form expressions can be found in Appendix A.

281 As opposed to the majority of state-of-the-art methods, the proposed algorithms  
 282 are also able to estimate the degradation matrices  $\mathbf{P}_i$  for  $i \in \{1, \dots, 3\}$ . These updates  
 283 can be derived based on the classical MU associated to the matrix model  $\mathbf{X} = \mathbf{U} \mathbf{V}^\top$   
 284 [9]. For  $\mathbf{P}_1$ , we are interested in solving  $D_\beta(\mathbf{Y}_1^{(1)} \|\mathbf{P}_1 \mathbf{A} (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B})^\top)$ . By posing  
 285  $\mathbf{V}^\top = \mathbf{A} (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B})^\top$ , we derive:

$$286 \quad (4.3) \quad \mathbf{P}_1 \leftarrow \tilde{\mathbf{P}}_1 \square \left( \frac{\left[ (\tilde{\mathbf{P}}_1 \mathbf{V}^\top)^{(\beta-2)} \square \mathbf{Y}_1^{(1)} \right] \mathbf{V}}{\left[ \tilde{\mathbf{P}}_1 \mathbf{V}^\top \right]^{\cdot (\beta-1)} \mathbf{V}} \right)^{\cdot \gamma(\beta)}.$$

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

289 Algorithm 4.1 summarizes the proposed method to tackle the  $\beta$ - $(L_r, L_r, 1)$ -NBTD  
 290 optimization problem (3.3). It consists in two optimization loops:

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

294 **Loop 2:** All the factors, including the matrices  $\mathbf{P}_i$ , are alternatively updated. The  
 295 maximum number of iterations for Loop 2 is `i2`. For the HSR problem, the operators  
 296  $\mathbf{P}_i$  for  $i \in \{1, \dots, 3\}$  are usually known and therefore the parameter `i2` is set to zero.  
 297 Loop 2 is considered in the case we have partial knowledge or uncertainties on one of  
 298 more downsampling operators, similarly as done in [25] with a matrix model. This  
 299 case will be later referred to as “semi-blind”.

300 The Algorithm is stopped when the relative change of the cost function  $\Phi$  from  
 301 (3.3) is below some given threshold  $\kappa$ , or when the maximum number of iterations is  
 302 reached.

303 **Initialization:** Many options are available to initialize factors  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ . In this  
 304 paper, an efficient way to initialize the low-rank factors of the  $L_r L_r 1$ -BTD is proposed.  
 305 For factor  $\mathbf{C}$ , VCA [31] is performed on the HSI  $\mathbf{Y}_1$ , thus extracting high-resolution  
 306 spectral information. Then, the matrix  $\mathbf{S}$  of vectorized abundance maps is obtained  
 307 by solving the following inverse problem under nonnegative constraints:

$$308 \quad (4.4) \quad \min_{\mathbf{S} \geq 0} \mathbf{P}_3 \mathbf{C} \mathbf{S}^\top \approx \mathbf{Y}_2^{(3)}.$$

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

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



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

**Require:** Input tensors  $\mathcal{Y}_1 \geq \mathbf{0}$ ,  $\mathcal{Y}_2 \geq \mathbf{0}$ , initializations  $\mathbf{A}, \mathbf{B}, \mathbf{C} \geq \mathbf{0}$ , downsampling operators  $\mathbf{P}_i \geq \mathbf{0}$  for  $i \in \{1, \dots, 3\}$ ;  $R$ , ranks  $\{L_r\}_{r=1}^R$ , maximum number of iterations  $\mathbf{i}1$  and  $\mathbf{i}2$ , a threshold  $0 < \kappa \ll 1$ , and a weight  $\lambda > 0$ .

**Ensure:** An approximate solution to (3.3) under constraints (3.2)

```

1: % Loop 1: update of matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  only
2:  $i \leftarrow 0$ ,  $\Phi^0 = 1$ ,  $\Phi^1 = 0$ .
3: while  $i < \mathbf{i}1$  and  $\left| \frac{\Phi^i - \Phi^{i+1}}{\Phi^i} \right| > \kappa$  do
4:   Update  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  sequentially; see Equations (4.2)
5:   Compute the objective function  $\Phi^{i+1}$ 
6: end while
7: % Loop 2 : update of  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{P}_i$ 
8:  $i \leftarrow 0$ 
9: while  $i < \mathbf{i}2$  and  $\left| \frac{\Phi^i - \Phi^{i+1}}{\Phi^i} \right| > \kappa$  do
10:  Update  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{P}_i$  for  $i \in \{1, \dots, 3\}$  sequentially; see Equations (4.2)
    and (4.3)
11:  Compute the objective function  $\Phi^{i+1}$ 
12: end while
13: return  $\hat{\mathcal{Y}} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{c}_r$ 

```

**Algorithm 4.2** Initialization of Algorithm 4.1

**Require:** Input tensors  $\mathcal{Y}_1 \geq \mathbf{0}$ ,  $\mathcal{Y}_2 \geq \mathbf{0}$ , downsampling operators  $\mathbf{P}_i \geq \mathbf{0}$  for  $i \in \{1, \dots, 3\}$ ;  $R$ , ranks  $\{L_r\}_{r=1}^R$ , maximum number of iterations  $\mathbf{j}1$  and  $\mathbf{j}2$ , a threshold  $0 < \kappa \ll 1$ .

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

```

1: Initialize  $\mathbf{C}$  using VCA on  $\mathcal{Y}_1$ ;
2: Compute  $\mathbf{S}$  using result of (4.4);
3: % Non-blind case: update of  $\mathbf{A}_r$ ,  $\mathbf{B}_r$  only
4:  $j \leftarrow 0$ ,  $\Phi^0 = 1$ ,  $\Phi^1 = 0$ .
5: while  $j < \mathbf{j}1$  and  $\left| \frac{\Phi^j - \Phi^{j+1}}{\Phi^j} \right| > \kappa$  do
6:   Update  $\mathbf{A}_r$ ,  $\mathbf{B}_r$  using the classical MU
7:   Compute the objective function  $\Phi^{j+1}$ 
8: end while
9: % Blind case: Update  $\mathbf{A}_r$ ,  $\mathbf{B}_r$ , and  $\mathbf{P}_i$  for  $i \in \{1, \dots, 3\}$ 
10:  $j \leftarrow 0$ 
11: while  $j < \mathbf{j}2$  and  $\left| \frac{\Phi^j - \Phi^{j+1}}{\Phi^j} \right| > \kappa$  do
12:  Update  $\mathbf{A}_r$ ,  $\mathbf{B}_r$ , and  $\mathbf{P}_i$  using matrix-based MU
13:  Compute the objective function  $\Phi^{j+1}$ 
14: end while
15: return  $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_R]$ ,  $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_R]$  and  $\mathbf{C}$ 

```

314 **4.2. Nonnegative minimum-volume.** The primary approach for addressing  
315 the “min-vol  $\text{KL-}(L_r, L_r, 1)$ -NBTD” problem in equation (3.5) is analogous to the  
316 methodology employed in Section 4.1, specifically the cyclic block majorization mini-  
317 mization (BMM) framework. This framework involves cyclically updating each block  
318 of variables with the others kept fixed, with each update accomplished by minimizing  
319 a majorizer constructed using the current iterate. The MU for updating the factors

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

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

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

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

$$334 \quad \log \det(\mathbf{Q}) \leq \log \det(\tilde{\mathbf{Q}}) + \langle \tilde{\mathbf{Q}}^{-1}, \mathbf{Q} - \tilde{\mathbf{Q}} \rangle = \langle \tilde{\mathbf{Q}}^{-1}, \mathbf{Q} \rangle + \text{cst},$$

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

$$337 \quad (4.6) \quad \log \det(\mathbf{C}^T\mathbf{C} + \delta\mathbf{I}) \leq \langle \tilde{\mathbf{Q}}^{-1}, \mathbf{C}^T\mathbf{C} \rangle + \text{cst} = \text{trace}(\mathbf{C}\tilde{\mathbf{Q}}^{-1}\mathbf{C}^T) + \text{cst},$$

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

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

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

$$349 \quad (4.8) \quad \mathbf{C}^*(\boldsymbol{\mu}) = \tilde{\mathbf{C}} \odot \frac{\left[ \left[ [\mathbf{Q} + \mathbf{e}\boldsymbol{\mu}^T]^{-2} + \mathbf{R} \right]^{\frac{1}{2}} - (\mathbf{Q} + \mathbf{e}\boldsymbol{\mu}^T) \right]}{[\mathbf{D}]},$$

350 where  $\mathbf{Q} = \mathbf{E}_{K,I_1 J_1} \mathbf{S}^T - 4\gamma(\tilde{\mathbf{C}}\mathbf{U}^-) + \lambda \mathbf{P}_3^T \mathbf{E}_{K_2, IJ} \mathbf{H}^T$ ,  $\mathbf{D} = 4\gamma\tilde{\mathbf{C}}(\mathbf{U}^+ + \mathbf{U}^-)$ , and  
 351  $\mathbf{R} = 8\gamma\tilde{\mathbf{C}}(\mathbf{U}^+ + \mathbf{U}^-) \odot \left( \begin{bmatrix} \mathbf{Y}_1^{(3)} \\ \mathbf{C}\mathbf{H}_1 \end{bmatrix} \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \begin{bmatrix} \mathbf{Y}_2^{(3)} \\ \mathbf{P}_3\mathbf{C}\mathbf{H}_2 \end{bmatrix} \mathbf{H}_2^T \right)$  with  $\mathbf{U} = \mathbf{U}^+ - \mathbf{U}^- =$

<sup>1</sup> tight at the current iterate  $\tilde{\mathbf{C}}$ , that is  $\sum_k^K \bar{l}(\tilde{\mathbf{c}}_k|\tilde{\mathbf{c}}_k) = \text{trace}(\tilde{\mathbf{C}}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{C}}^T)$

352  $(\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} + \delta \mathbf{I})^{-1}$ ,  $\mathbf{U}^+ = \max(\mathbf{U}, \mathbf{0}) \geq \mathbf{0}$  and  $\mathbf{U}^- = \max(-\mathbf{U}, \mathbf{0}) \geq \mathbf{0}$ , and  $\mathbf{E}_{K, I_1, J_1}$  is  
 353 the  $K$ -by- $I_1 J_1$  matrix of all ones. As proved in [26, Proposition. 2], the constraint  
 354  $e^T \mathbf{C}^*(\mu) = e^T$  is satisfied for a unique  $\mu \in \mathbb{R}^R$ . Therefore a Newton-Raphson method  
 355 can be used to find the  $\mu$  with quadratic rate of convergence; see [26, Proposition. 3].  
 356 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  $\mathcal{Y}_1 \geq \mathbf{0}$ ,  $\mathcal{Y}_2 \geq \mathbf{0}$ , initializations  $\mathbf{A}, \mathbf{B}, \mathbf{C} \geq \mathbf{0}$ , downsampling operators  $\mathbf{P}_i \geq \mathbf{0}$  for  $i \in \{1, \dots, 3\}$ ;  $R$ , 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)

```

1: % Loop 1
2:  $i \leftarrow 0$ ,  $\Phi^0 = 1$ ,  $\Phi^1 = 0$ .
3: while  $i < i1$  and  $\left| \frac{\Phi^i - \Phi^{i+1}}{\Phi^i} \right| > \kappa$  do
4:   % Update of matrices  $\mathbf{A}$ ,  $\mathbf{B}$ 
5:   Update  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  sequentially; see Equations (4.2)
6:   % Update of matrix  $\mathbf{C}$ 
7:    $\mathbf{U} \leftarrow (\mathbf{C}^T \mathbf{C} + \delta \mathbf{I})^{-1}$ 
8:    $\mathbf{U}^+ \leftarrow \max(\mathbf{U}, \mathbf{0})$ 
9:    $\mathbf{U}^- \leftarrow \max(-\mathbf{U}, \mathbf{0})$ 
10:   $\mathbf{Q} \leftarrow \mathbf{E}_{K, I_1, J_1} \mathbf{H}_1^T - 4\gamma(\mathbf{C}\mathbf{U}^-) + \lambda \mathbf{P}_3^T \mathbf{E}_{K_2, IJ} \mathbf{H}_2^T$ 
11:   $\mathbf{D} \leftarrow 4\gamma \mathbf{C}(\mathbf{U}^+ + \mathbf{U}^-)$ 
12:   $\mathbf{R} \leftarrow 8\gamma \mathbf{C}(\mathbf{U}^+ + \mathbf{U}^-) \odot \left( \begin{bmatrix} \mathbf{Y}_1^{(3)} \\ \mathbf{C}\mathbf{H}_1 \end{bmatrix} \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \begin{bmatrix} \mathbf{Y}_2^{(3)} \\ \mathbf{P}_3 \mathbf{C}\mathbf{H}_2 \end{bmatrix} \mathbf{H}_2^T \right)$ 
13:   $\mu \leftarrow \text{root}(e^T \mathbf{C}^*(\mu) = e^T)$  over  $\mathbb{R}^R$  % see (4.8) for the expression of  $\mathbf{C}^*(\mu)$ 
14:   $\mathbf{C} \leftarrow \mathbf{C} \odot \frac{\left[ \left[ (\mathbf{Q} + e\mu^T)^2 + \mathbf{R} \right]^{\frac{1}{2}} - (\mathbf{Q} + e\mu^T) \right]}{[\mathbf{D}]}$ 
15:  Compute the objective function  $\Phi^{i+1}$ 
16: end while
17: % Loop 2
18:  $i \leftarrow 0$ 
19: while  $i < i2$  and  $\left| \frac{\Phi^i - \Phi^{i+1}}{\Phi^i} \right| > \kappa$  do
20:   Update  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{P}_i$  for  $i \in \{1, \dots, 3\}$  sequentially; see Equations (4.2),
    (4.3) and update of  $\mathbf{C}$  as performed in Loop 1
21:   Compute the objective function  $\Phi^{i+1}$ 
22: end while
23: return  $\hat{\mathcal{Y}} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \otimes \mathbf{c}_r$ 

```

---

357 **4.3. Key insights about the Algorithms.** The following paragraphs discuss  
 358 key elements for Algorithms 4.1 and 4.3 proposed in previous sections.

359 **Comments on computational complexity:** The computational complexity of Al-  
 360 gorithm 4.1, designed to solve the  $\beta$ - $(L_r, L_r, 1)$ -NBT problem, exhibits asymptotic  
 361 equivalence to the standard MU for  $\beta$ -NMF, after unfoldings. This implies that it re-  
 362 quires  $\mathcal{O}(IJK \times \sum_r^R L_r)$  operations per iteration. The main driver of this complexity  
 363 lies in the aforementioned matrix products. One might question the influence of the  
 364 computation step of Algorithm 4.3, which aims to calculate the optimal Lagrangian  
 365 multipliers, thereby ensuring the column-stochasticity constraints of matrix  $\mathbf{C}$  are sat-  
 366 isfied. In [26], the authors demonstrate the quadratic convergence of the procedure

367 employed to compute the Lagrangian multipliers. As a result, the number of iterations  
 368 required for this process remains low in practical scenarios. Consequently, the overall  
 369 complexity of Algorithm 4.3 is predominantly determined by the computationally in-  
 370 tensive matrix products mentioned earlier. For more comprehensive information, we  
 371 refer the reader to [26].

372 **Parallelization:** Some of the most computationally intensive steps of the proposed  
 373 algorithm can be easily ran onto a parallel computation platform. Indeed, the com-  
 374 plexity of the MU given in Equation (4.2), for instance, is mainly driven by the matrix  
 375 products in which matrices  $\mathbf{A}$  and  $\mathbf{H}$  are involved. On MATLAB for example, one  
 376 can easily take of advantage of a GPU compatible with CUDA libraries by simply  
 377 transforming usual arrays into GPU arrays and significantly speed up the Algorithm.

378 **Comments on convergence guarantees:** In practice, the updates of factors  $\mathbf{A}$ ,  $\mathbf{B}$ ,  
 379  $\mathbf{C}$  and  $\mathbf{P}_i$  are obtained from the element-wise maximum between the matrix updates,  
 380 that correspond to the closed form expression of the minimizer of the majorization  
 381 built at the current iterate [9, 25], and a small positive scalar  $\epsilon$  (here we choose the  
 382 Matlab machine epsilon). These modified updates aim at establishing convergence  
 383 guarantee to stationary points within the Block Successive Minimization Methods  
 384 (BSUM) framework [37].

385 **5. Numerical experiments.** All tests are preformed using Matlab R2021a on  
 386 a laptop Intel CORE i7-11800H CPU @2.30GHz 16GB RAM with GeForce RTX3060  
 387 GPU. The code is available online at [https://github.com/cprevost4/bLL1\\_NBTD](https://github.com/cprevost4/bLL1_NBTD).

388 **5.1. Test setup.** The proposed algorithms were compared to several tensor  
 389 methods designed for solving the HSR problem, namely STEREO and Blind-STEREO  
 390 [17, 19], SCOTT and BSCOTT [34], CT-STAR and CB-STAR [2], SCLL1 [6] and  
 391 CNN-BTD-Var [32]. Among them, SCLL1 and CNN-BTD-Var were based on the  
 392 LL1-BTD model<sup>2</sup>, therefore they are able to solve the unmixing problem<sup>3</sup>.

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

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

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

$$405 \quad (5.1) \quad \text{PSNR} = 10 \log_{10} \left( \frac{\|\mathcal{Y}\|_F^2}{\|\hat{\mathcal{Y}} - \mathcal{Y}\|_F^2} \right).$$

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

$$407 \quad (5.2) \quad \text{CC} = \frac{1}{IJK} \left( \sum_{k=1}^K \rho(\mathbf{y}_{:, :, k}, \hat{\mathbf{y}}_{:, :, k}) \right),$$

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

<sup>2</sup>These works all considered the simpler model with  $L_1 = \dots = L_R$ .

<sup>3</sup>The performance of SCLL1 for blind spectral unmixing was not assessed in the original work.

409 spectral slices;

$$410 \quad (5.3) \quad \text{ERGAS} = \frac{100}{d} \sqrt{\frac{1}{IJK} \sum_{k=1}^K \frac{\|\hat{\mathbf{Y}}_{::,k} - \mathbf{Y}_{::,k}\|_F^2}{\mu_k^2}},$$

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

$$414 \quad (5.4) \quad \text{SAD} = \frac{1}{R} \sum_{r=1}^R \arccos \left( \frac{\mathbf{c}_r^\top \hat{\mathbf{c}}_r}{\|\mathbf{c}_r\|_2 \|\hat{\mathbf{c}}_r\|_2} \right),$$

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

$$419 \quad (5.5) \quad \text{RMSE} = \frac{1}{R} \sum_{r=1}^R \sqrt{\frac{1}{IJ} \sum_{d=1}^{IJ} \left( (\mathbf{S}_r)_d - (\hat{\mathbf{S}}_r)_d \right)^2}.$$

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

425 **5.2. Degradation model.** We considered synthetic and semi-real datasets for  
 426 which a reference SRI is available. The HSI was obtained by spatial degradation of  
 427  $\mathbf{Y}$  using  $\mathbf{P}_1$  and  $\mathbf{P}_2$  while the MSI was obtained by spectral degradation of  $\mathbf{Y}$  with  
 428  $\mathbf{P}_3$  according to model (2.5).

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

433 For the spectral degradation matrix  $\mathbf{P}_3$ , the spectral response functions of the  
 434 Sentinel-2 instrument<sup>4</sup> was used. It spans the electromagnetic spectrum from 412nm  
 435 to 2022nm and produces a 10-band MSI ( $K_2 = 10$ ) corresponding to the wavelengths  
 436 433–453nm (atmospheric correction), 458–522nm (soil, vegetation), 543–577nm (green  
 437 peak), 650–680nm (maximum chlorophyll absorption), 698–712nm (red edge), 733–  
 438 747nm (red edge), 773–793nm (leaf area index, edge of NIR), 785–900nm (leaf area  
 439 index), 855–875nm (NIR plateau), 935–955nm (water vapour absorption). The spec-  
 440 tral degradation matrix  $\mathbf{P}_3$  acts as a selection-weighting matrix, utilizing the shared  
 441 spectral bands between the SRI and the MSI datasets.

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

<sup>4</sup>available for download at <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/assetpublisher/Wk0TKajiSaR/content/sentinel-2a-spectral-responses>.

447 **5.3. Datasets.** Experiments were run on synthetic and semi-real datasets de-  
 448 scribed below.

449 **Synthetic dataset:** A dataset where the SRI admits an exact  $L_r L_r 1$ -BTD was  
 450 created, in order to perform simulations in a controlled environment.

451 We considered  $R = 4$  spectral signatures  $\mathbf{c}_r$  ( $r \in \{1, \dots, R\}$ ) obtained from the  
 452 Jasper Ridge reference data<sup>5</sup>. The SRI  $\mathbf{Y} \in \mathbb{R}^{I \times J \times K}$  ( $I = J = 120, K = 173$ ) was split  
 453 into 36 equal blocks in the spatial dimensions. We set  $L_1 = L_2 = 3$  and  $L_3 = L_4 = 6$ .  
 454 Each abundance map  $\mathbf{S}_r$  ( $r \in \{1, \dots, R\}$ ) was a block matrix with  $L_r$  blocks of size  
 455  $\frac{I}{L_r} \times \frac{J}{L_r}$ . Each block in the parcel map was a patch composed of entries equal to  
 456 one. At most one material is active in each block, thus the so-called “pure pixel  
 457 assumption” was valid. Figure 1 depicts the abundance maps for the four materials  
 in the synthetic data set.

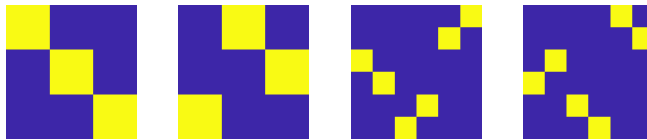


FIG. 1. Abundance maps for the four materials in the synthetic dataset.

458 The reference SRI is computed as follows:  
 459

$$460 \quad \mathbf{Y} = \sum_{r=1}^R \mathbf{S}_r \otimes \mathbf{c}_r.$$

461

462 **Semi-real datasets:** The first semi-real dataset we considered is built upon the  
 463 Jasper Ridge reference SRI  $\mathbf{Y} \in \mathbb{R}^{100 \times 100 \times 173}$ . It includes four materials: road, soil,  
 464 water and vegetation. We chose  $R = 4$ , and  $L_1 = 15, L_2 = 8, L_3 = 20, L_4 = 13$ .

465 A second semi-real dataset was considered and based on the Ivanpah Playa [30]  
 466 reference SRI  $\mathbf{Y} \in \mathbb{R}^{95 \times 95 \times 156}$ . This dataset is composed of four, very similar materials:  
 467 road, solar panels, light sand and darker sand. Thus it constitutes a challenging  
 468 example for which unmixing may lead to rank-deficient solutions.

469 **5.4. Reconstruction results.** In order to assess the performance of our ap-  
 470 proaches for data fusion, the quality metrics obtained with each method over 5 trials  
 471 are reported in the Tables 1–3 below. The two best metrics of each column are  
 472 shown in bold. For comparison with Algorithm 4.3, we only show the results for the  
 473 KL divergence, *i.e.*, for  $\beta = 1$  along with Poisson noise. The results for other noise  
 474 statistics are available in supplementary materials.

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

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

485 **5.5. Unmixing results.** For the unmixing task, the reference and estimated  
 486 spectral signatures and abundance maps are depicted. Due to space limitations,

<sup>5</sup>Available for download at <http://lesun.weebly.com/hyperspectral-data-set.html>.

Method	CC ( $\uparrow$ )	SAD ( $\downarrow$ )	RMSE ( $\downarrow$ )	ERGAS ( $\downarrow$ )	PSNR (dB) ( $\uparrow$ )	Time (sec) ( $\downarrow$ )	C.R. ( $\uparrow$ )
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	<b>266</b>
CT-STAR	<b>0.999</b>	<b>0.25</b>	<b>5.272e-3</b>	<b>0.713</b>	<b>39.71</b>	<b>0.102</b>	242
CB-STAR	<b>0.999</b>	1.25	<b>6.408e-3</b>	0.991	<b>38.02</b>	20.14	138
CNN-BTD-Var	0.997	5.30	1.260e-2	2.952	32.76	3.302	266
CNMF	0.999	<b>1.09</b>	7.313e-3	1.364	36.20	1.872	43
FUSE	0.994	2.77	1.470e-2	2.813	30.92	<b>0.146</b>	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	<b>0.776</b>	35.69	23.42	<b>497</b>
Alg. 4.3 ( $\beta = 1$ )	0.999	2.96	8.703e-3	2.007	36.88	30.19	<b>497</b>

TABLE 1  
Reconstruction of the synthetic dataset with Poisson noise.

Method	CC ( $\uparrow$ )	SAD ( $\downarrow$ )	RMSE ( $\downarrow$ )	ERGAS ( $\downarrow$ )	PSNR (dB) ( $\uparrow$ )	Time (sec) ( $\downarrow$ )	C.R. ( $\uparrow$ )
STEREO	0.989	3.01	1.378e-2	2.617	<b>31.10</b>	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	<b>0.180</b>	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	<b>0.092</b>	<b>456</b>
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	<b>199</b>
CNMF	<b>0.999</b>	<b>0.90</b>	<b>6.000e-3</b>	<b>1.493</b>	27.96	1.302	42
FUSE	0.984	2.88	2.066e-2	3.227	28.15	0.276	8
HySure	<b>0.998</b>	<b>2.05</b>	<b>1.196e-2</b>	<b>1.778</b>	<b>33.55</b>	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.

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

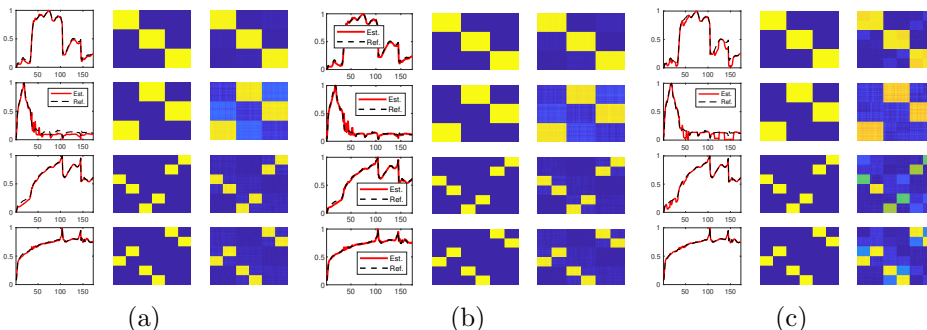


FIG. 2. Unmixing on the synthetic dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.

Method	CC ( $\uparrow$ )	SAD ( $\downarrow$ )	RMSE ( $\downarrow$ )	ERGAS ( $\downarrow$ )	PSNR (dB) ( $\uparrow$ )	Time (sec) ( $\downarrow$ )	C.R. ( $\uparrow$ )
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	<b>196</b>
BSCOTT	0.870	5.45	3.946e-2	3.320	23.82	<b>0.115</b>	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	<b>0.053</b>	<b>442</b>
CB-STAR	0.988	1.24	1.215e-2	0.969	32.31	5.955	196
CNN-BTD-Var	0.973	<b>1.11</b>	1.827e-2	1.295	28.42	2.727	86
CNMF	<b>0.988</b>	1.12	<b>8.598e-3</b>	<b>0.766</b>	<b>34.96</b>	1.931	42
FUSE	<b>0.991</b>	<b>1.05</b>	<b>9.621e-3</b>	<b>0.825</b>	<b>34.02</b>	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.

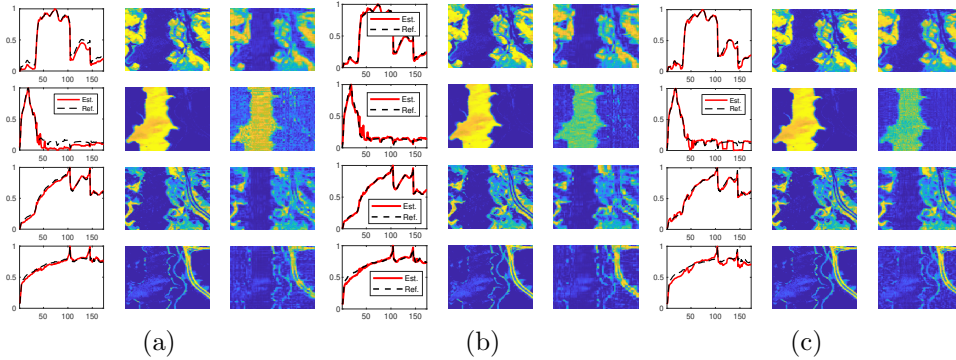


FIG. 3. Unmixing on the Jasper Ridge dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.

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

499 **5.6. Performance in a semi-blind scenario.** The performance of our algo-  
 500 rithms was then evaluated in a semi-blind scenario, *i.e.*, when the spatial degradation  
 501 matrices  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are unknown. This scenario is likely to occur, e.g., when the  
 502 HSI and MSI are acquired at different times or in the presence of motion blur. The  
 503 proposed algorithms were compared to two semi-blind tensor algorithms: BSTEREO  
 504 and BSCOTT, that do not estimate  $\mathbf{P}_1$  and  $\mathbf{P}_2$ .

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

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



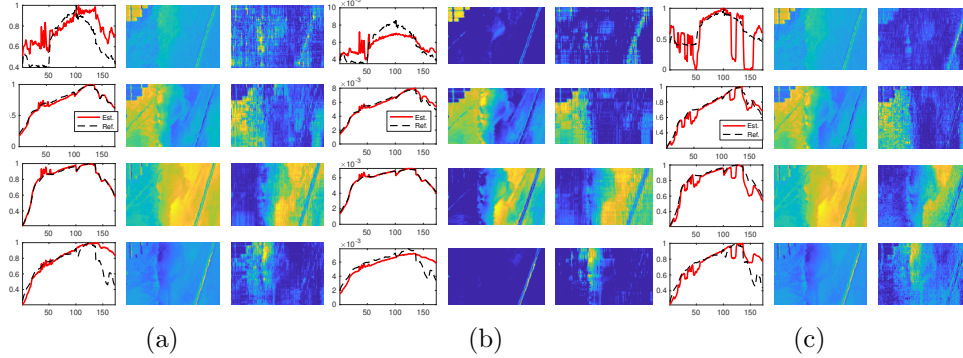


FIG. 4. Unmixing on the Ivanpah Playa dataset with Poisson noise, (a) Algorithm 4.1, (b) Algorithm 4.3, (c) SCLL1.

Method	CC ( $\uparrow$ )	SAD ( $\downarrow$ )	RMSE ( $\downarrow$ )	ERGAS ( $\downarrow$ )	PSNR (dB) ( $\uparrow$ )	Time (sec) ( $\downarrow$ )
<b>Gaussian Noise - 30 dB</b>						
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	<b>0.108</b>
MR- $\beta$ -NMF	0.971	6.88	4.508e-2	6.120	22.34	1393.747
Alg. 4.1 ( $\beta = 2$ )	<b>0.976</b>	<b>3.85</b>	<b>2.521e-2</b>	<b>4.407</b>	<b>26.36</b>	74.142
<b>Poisson Noise - 30dB</b>						
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	<b>0.988</b>	4.57	2.306e-2	3.228	<b>28.83</b>	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	<b>4.34</b>	<b>2.161e-2</b>	<b>2.937</b>	27.76	26.59
<b>Gamma Noise</b>						
BSTEREO	0.961	5.57	<b>2.691e-2</b>	4.503	24.82	1.689
BSCOTT	0.951	5.80	3.258e-2	4.936	24.13	<b>0.483</b>
MR- $\beta$ -NMF	<b>0.981</b>	5.41	3.335e-2	<b>4.46</b>	25.58	1401.120
Alg. 4.1 ( $\beta = 0$ )	0.967	<b>3.66</b>	3.024e-2	5.627	<b>25.93</b>	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-

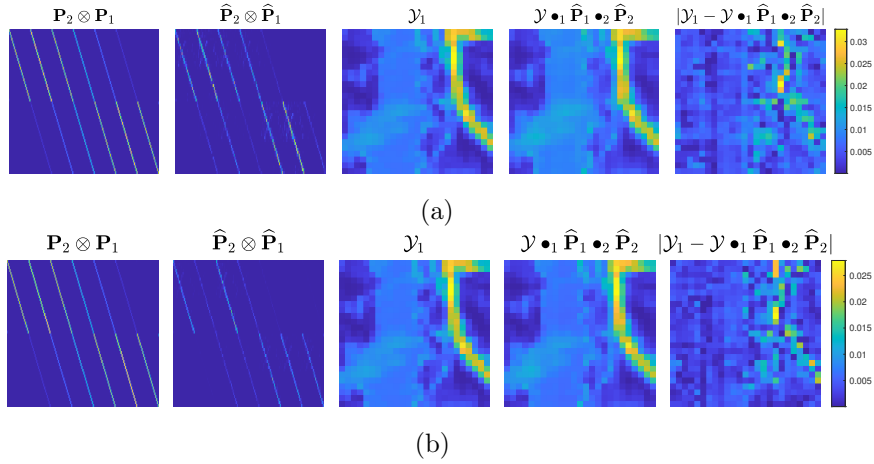


FIG. 5. Estimation of the spatial degradation and comparison with the real HSI, (a) Algorithm 4.1 ( $\beta = 1$ ), (b) Algorithm 4.3 ( $\beta = 1$ ).

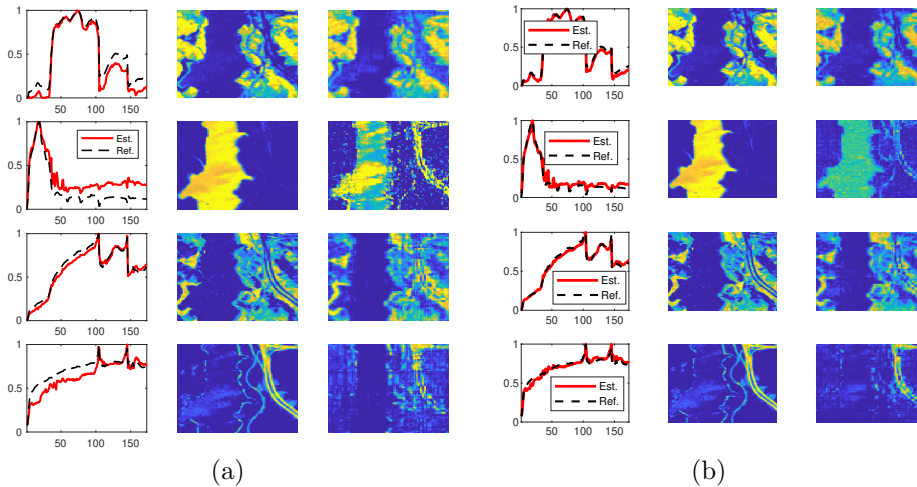


FIG. 6. Semi-blind unmixing, (a) Algorithm 4.1 ( $\beta = 1$ ), (b) Algorithm 4.3 ( $\beta = 1$ ).

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

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

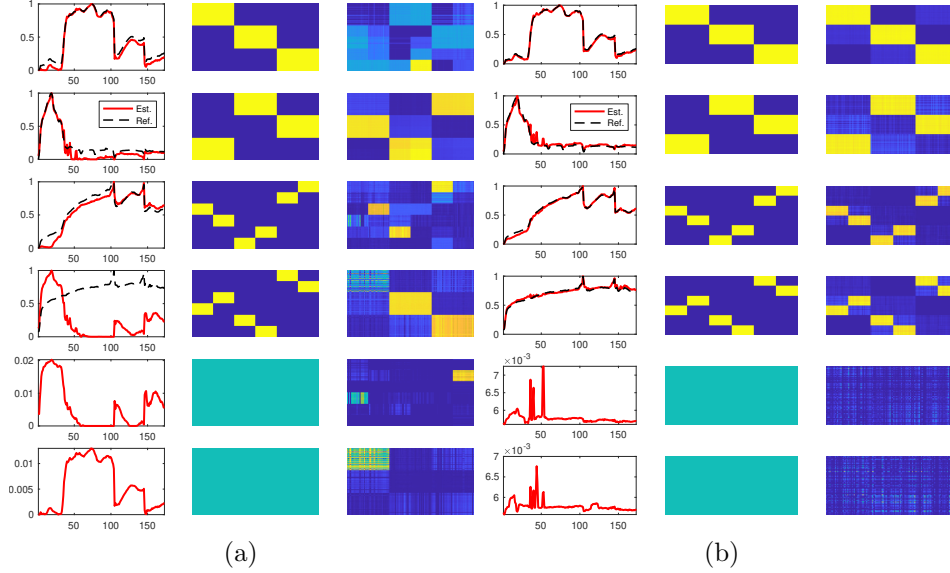


FIG. 7. Unmixing results with overestimation of the rank  $R$ , (a) Algorithm 4.1 ( $\beta = 1$ ), (b) Algorithm 4.3 ( $\beta = 1$ ).

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

553 **Appendix A. Detailed MU updates.** This appendix contains the closed-  
 554 form expressions for the updates of  $\mathbf{B}$  and  $\mathbf{C}$ , similarly to (4.2).  
 (A.1)

$$555 \quad \mathbf{B} = \tilde{\mathbf{B}} \square \left( \frac{\left[ \mathbf{P}_2^T \left( \left( \mathbf{P}_2 \tilde{\mathbf{B}} \mathbf{H}_1 \right)^{(\beta-2)} \square \mathbf{Y}_1^{(2)} \right) \mathbf{H}_1^T + \lambda \left( \left( \tilde{\mathbf{B}} \mathbf{H}_2 \right)^{(\beta-2)} \square \mathbf{Y}_2^{(2)} \right) \mathbf{H}_2^T \right]^{\gamma(\beta)}}{\left[ \mathbf{P}_2^T \left( \mathbf{P}_2 \tilde{\mathbf{B}} \mathbf{H}_1 \right)^{(\beta-1)} \mathbf{H}_1^T + \lambda \left( \tilde{\mathbf{B}} \mathbf{H}_2 \right)^{(\beta-1)} \mathbf{H}_2^T \right]} \right),$$

556 where  $\mathbf{H}_1 = (\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A})^\top$  and  $\mathbf{H}_2 = (\mathbf{P}_3 \mathbf{C} \odot_p \mathbf{A})^\top$ . For  $\mathbf{C}$ , we have  
 (A.2)

$$557 \quad \mathbf{C} = \tilde{\mathbf{C}} \square \left( \frac{\left[ \left( \left( \tilde{\mathbf{C}} \mathbf{H}_1 \right)^{(\beta-2)} \square \mathbf{Y}_1^{(3)} \right) \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \left( \left( \mathbf{P}_3 \tilde{\mathbf{C}} \mathbf{H}_2 \right)^{(\beta-2)} \square \mathbf{Y}_2^{(3)} \right) \mathbf{H}_2^T \right]^{\gamma(\beta)}}{\left[ \left( \tilde{\mathbf{C}} \mathbf{H}_1 \right)^{(\beta-1)} \mathbf{H}_1^T + \lambda \mathbf{P}_3^T \left( \mathbf{P}_3 \tilde{\mathbf{C}} \mathbf{H}_2 \right)^{(\beta-1)} \mathbf{H}_2^T \right]} \right),$$

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

559 The updates for  $\mathbf{P}_2$  (resp.  $\mathbf{P}_3$ ) are obtained by substituting  $\mathbf{P}_1$  by  $\mathbf{P}_2$  (resp.  
 560  $\mathbf{P}_3$ ),  $\mathbf{Y}_1^{(1)}$  by  $\mathbf{Y}_1^{(2)}$  (resp.  $\mathbf{Y}_2^{(3)}$ ) and defining  $\mathbf{V}^\top = \mathbf{B}(\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A})^\top$  (resp.  $\mathbf{V}^\top =$   
 561  $\mathbf{C}(\mathbf{A} \odot_{\text{vec}} \mathbf{B})^\top$ ) in (4.3).

- 563 [1] J. M. BIOUCAS-DIAS AND J. P. NASCIMENTO, *Hyperspectral subspace identification*, IEEE Trans.  
564 Geosci. Remote Sens., 46 (2008), pp. 2435–2445.
- 565 [2] R. A. BORSOI, C. PRÉVOST, K. USEVICH, D. BRIE, J. M. BERMUDEZ, AND C. RICHARD, *Coupled*  
566 *tensor decomposition for hyperspectral and multispectral image fusion with inter-image*  
567 *variability*, IEEE Journal of Selected Topics in Signal Processing, 15 (2021), pp. 702–717,  
568 <https://doi.org/10.1109/JSTSP.2021.3054338>.
- 569 [3] M. BOUSSE, O. DEBALS, AND L. DE LATHAUWER, *A tensor-based method for large-scale blind*  
570 *source separation using segmentation*, IEEE Trans. Signal Process., 65 (2016), pp. 346–358.
- 571 [4] R. BRO AND N. D. SIDIROPOULOS, *Least squares algorithms under unimodality and non-*  
572 *negativity constraints*, Journal of Chemometrics: A Journal of the Chemometrics Society,  
573 12 (1998), pp. 223–247.
- 574 [5] P. COMON, *Tensors: A brief introduction*, IEEE Signal Process. Mag., 31 (2014), pp. 44–53.
- 575 [6] M. DING, X. FU, T.-Z. HUANG, J. WANG, AND X.-L. ZHAO, *Hyperspectral super-resolution*  
576 *via interpretable block-term tensor modeling*, arXiv e-prints, (2020), arXiv:2006.10248,  
577 p. arXiv:2006.10248, <https://arxiv.org/abs/2006.10248>.
- 578 [7] C. FÉVOTTE, N. BERTIN, AND J.-L. DURRIEU, *Nonnegative matrix factorization with the*  
579 *Itakura-Saito divergence: With application to music analysis*, Neural computation, 21  
580 (2009), pp. 793–830.
- 581 [8] C. FÉVOTTE AND N. DOBIGEON, *Nonlinear hyperspectral unmixing with robust nonnegative*  
582 *matrix factorization*, IEEE Transactions on Image Processing, 24 (2015), pp. 4810–4819.
- 583 [9] C. FÉVOTTE AND J. IDIER, *Algorithms for nonnegative matrix factorization with the  $\beta$ -*  
584 *divergence*, Neural computation, 23 (2011), pp. 2421–2456.
- 585 [10] X. FU, K. HUANG, N. D. SIDIROPOULOS, AND W.-K. MA, *Nonnegative matrix factorization*  
586 *for signal and data analytics: Identifiability, algorithms, and applications.*, IEEE Signal  
587 Process. Mag., 36 (2019), pp. 59–80.
- 588 [11] X. FU, W.-K. MA, T.-H. CHAN, AND J. M. BIOUCAS-DIAS, *Self-dictionary sparse regression*  
589 *for hyperspectral unmixing: Greedy pursuit and pure pixel search are related*, IEEE J. Sel.  
590 Topics Signal Process., 9 (2015), pp. 1128–1141.
- 591 [12] P. V. GIAMPOURAS, A. A. RONTOGIANNIS, AND E. KOFIDIS, *Block-term tensor decomposition*  
592 *model selection and computation: The bayesian way*, IEEE Transactions on Signal Process-  
593 ing, 70 (2022), pp. 1704–1717.
- 594 [13] N. GILLIS, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, 2020, [https://doi.org/10.](https://doi.org/10.1137/1.9781611976410)  
595 [1137/1.9781611976410](https://doi.org/10.1137/1.9781611976410).
- 596 [14] H. GUO, W. BAO, K. QU, X. MA, AND M. CAO, *Multispectral and hyperspectral image fu-*  
597 *sion based on regularized coupled non-negative block-term tensor decomposition*, Remote  
598 Sensing, 14 (2022), p. 5306.
- 599 [15] K. HUANG, N. D. SIDIROPOULOS, AND A. P. LIAVAS, *A flexible and efficient algorithmic frame-*  
600 *work for constrained matrix and tensor factorization*, IEEE Trans. Signal Process., 64  
601 (2016), pp. 5052–5065.
- 602 [16] T. IMBIRIBA, R. A. BORSOI, AND J. M. BERMUDEZ, *Generalized linear mixing model accounting*  
603 *for endmember variability*, in 2018 IEEE ICASSP, 2018, pp. 1862–1866.
- 604 [17] C. I. KANATSOU LIS, X. FU, N. D. SIDIROPOULOS, AND W.-K. MA, *Hyperspectral Super-*  
605 *Resolution: A Coupled Tensor Factorization Approach*, IEEE Trans. Signal Process., 66  
606 (2018), pp. 6503–6517.
- 607 [18] C. I. KANATSOU LIS, X. FU, N. D. SIDIROPOULOS, AND W.-K. MA, *Hyperspectral Super-*  
608 *Resolution: Combining Low Rank Tensor and Matrix Structure*, in 2018 IEEE ICIP, Oct.  
609 2018, pp. 3318–3322.
- 610 [19] C. I. KANATSOU LIS, X. FU, N. D. SIDIROPOULOS, AND W.-K. MA, *Hyperspectral super-resolution*  
611 *via coupled tensor factorization: Identifiability and algorithms*, in 2018 IEEE International  
612 Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 3191–3195.
- 613 [20] C. I. KANATSOU LIS AND N. D. SIDIROPOULOS, *Tex-graph: Coupled tensor-matrix knowledge-*  
614 *graph embedding for covid-19 drug repurposing*, in Proceedings of the 2021 SIAM Interna-  
615 tional Conference on Data Mining (SDM), SIAM, 2021, pp. 603–611.
- 616 [21] B. KING, C. FÉVOTTE, AND P. SMARAGDIS, *Optimal cost function and magnitude power for*  
617 *nmf-based speech separation and music interpolation*, in 2012 IEEE International Workshop  
618 on Machine Learning for Signal Processing, 2012, pp. 1–6, [https://doi.org/10.1109/MLSP.](https://doi.org/10.1109/MLSP.2012.6349726)  
619 [2012.6349726](https://doi.org/10.1109/MLSP.2012.6349726).
- 620 [22] T. G. KOLDA AND B. W. BADER, *Tensor Decompositions and Applications*, SIAM Review, 51  
621 (2009), pp. 455–500.
- 622 [23] S. M. KRIEG, N. H. BUCHMANN, J. GEMPT, E. SHIBAN, B. MEYER, AND F. RINGEL, *Diffusion*  
623 *tensor imaging fiber tracking using navigated brain stimulation—a feasibility study*, Acta  
624 neurochirurgica, 154 (2012), pp. 555–563.
- 625 [24] V. LEPLAT, N. GILLIS, AND A. M. ANG, *Blind audio source separation with minimum-volume*  
626 *beta-divergence nmf*, IEEE Transactions on Signal Processing, 68 (2020), pp. 3400–3410.
- 627 [25] V. LEPLAT, N. GILLIS, AND C. FÉVOTTE, *Multi-resolution beta-divergence nmf for blind spectral*  
628 *unmixing*, Signal Processing, (2021).
- 629 [26] V. LEPLAT, N. GILLIS, AND J. IDIER, *Multiplicative updates for nmf with  $\beta$ -divergences under*

- 630 disjoint equality constraints, *SIAM Journal on Matrix Analysis and Applications*, 42 (2021),  
 631 pp. 730–752.
- 632 [27] Q. LI, W.-K. MA, AND Q. WU, *Hyperspectral super-resolution: Exact recovery in polynomial*  
 633 *time*, in 2018 IEEE SSP, IEEE, 2018, pp. 378–382.
- 634 [28] C.-J. LIN, *Projected gradient methods for nonnegative matrix factorization*, *Neural computa-*  
 635 *tion*, 19 (2007), pp. 2756–2779.
- 636 [29] H. LIU, W. JIANG, Y. ZHA, AND Z. WEI, *Coupled tensor block term decomposition with*  
 637 *superpixel-based graph laplacian regularization for hyperspectral super-resolution*, *Remote*  
 638 *Sensing*, 14 (2022), p. 4520.
- 639 [30] S. MAHALINGAM, P. SRINIVAS, P. K. DEVI, D. SITA, S. K. DAS, T. S. LEELA, AND V. R.  
 640 VENKATARAMAN, *Reflectance based vicarious calibration of hysis sensors and spectral stab-*  
 641 *ility study over pseudo-invariant sites*, in 2019 IEEE Recent Advances in Geoscience and  
 642 Remote Sensing: Technologies, Standards and Applications (TENGARSS), IEEE, 2019,  
 643 pp. 132–136.
- 644 [31] J. NASCIMENTO AND J. DIAS, *Vertex component analysis: A fast algorithm to unmix hyper-*  
 645 *spectral data*, *IEEE Geosci. Remote Sens. Lett.*, 43 (2005), pp. 898–910.
- 646 [32] C. PRÉVOST, R. A. BORSOI, K. USEVICH, D. BRIE, J. M. BERMUDEZ, AND C. RICHARD, *Hy-*  
 647 *perspectral super-resolution accounting for spectral variability: Coupled tensor l11-based*  
 648 *recovery and blind unmixing of the unknown super-resolution image*, *SIAM Journal on*  
 649 *Imaging Sciences*, 15 (2022), pp. 110–138, <https://doi.org/10.1137/21M1409354>.
- 650 [33] C. PRÉVOST AND V. LEPLAT, *Nonnegative block-term decomposition with the  $\beta$ -divergence:*  
 651 *Joint data fusion and blind spectral unmixing*, in ICASSP 2023-2023 IEEE International  
 652 Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.
- 653 [34] C. PRÉVOST, K. USEVICH, P. COMON, AND D. BRIE, *Coupled tensor low-rank multilinear ap-*  
 654 *proximation for hyperspectral super-resolution*, in ICASSP 2019 - 2019 IEEE International  
 655 Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 5536–5540.
- 656 [35] C. PRÉVOST, K. USEVICH, P. COMON, AND D. BRIE, *Hyperspectral Super-Resolution with Cou-*  
 657 *pled Tucker Approximation: Identifiability and SVD-based algorithms*, *IEEE Trans. Signal*  
 658 *Process.*, 68 (2020), pp. 931–946.
- 659 [36] Y. QIAN, F. XIONG, S. ZENG, J. ZHOU, AND Y. TANG, *Matrix-vector nonnegative tensor fac-*  
 660 *torization for blind unmixing of hyperspectral imagery*, *IEEE Trans. Geosci. Remote Sens.*,  
 661 55 (2016), pp. 1776–1792.
- 662 [37] M. RAZAVIYAYN, M. HONG, AND Z.-Q. LUO, *A unified convergence analysis of block successive*  
 663 *minimization methods for nonsmooth optimization*, *SIAM Journal on Optimization*, 23  
 664 (2013), pp. 1126–1153.
- 665 [38] A. A. RONTOGIANNIS, E. KOFIDIS, AND P. V. GIAMPOURAS, *Online rank-revealing block-term*  
 666 *tensor decomposition*, *Signal Processing*, (2023), p. 109126.
- 667 [39] G. A. SHAW AND H. K. BURKE, *Spectral imaging for remote sensing*, *Lincoln laboratory journal*,  
 668 14 (2003), pp. 3–28.
- 669 [40] M. SIMOES, J. M. BIOUCAS-DIAS, L. B. ALMEIDA, AND J. CHANUSSOT, *A convex formulation for*  
 670 *hyperspectral image superresolution via subspace-based regularization*, *IEEE Trans. Geosci.*  
 671 *Remote Sens.*, 53 (2015), pp. 3373–3388.
- 672 [41] Y. SUN, P. BABU, AND D. PALOMAR, *Majorization-minimization algorithms in signal process-*  
 673 *ing, communications, and machine learning*, *IEEE Transactions on Signal Processing*, 65  
 674 (2017), pp. 794–816.
- 675 [42] L. WALD, T. RANCHIN, AND M. MANGOLINI, *Fusion of satellite images of different spatial*  
 676 *resolutions: Assessing the quality of resulting images*, *Photogrammetric Eng. and Remote*  
 677 *Sens.*, 63 (1997), pp. 691–699.
- 678 [43] M. WANG, D. HONG, Z. HAN, J. LI, J. YAO, L. G., B. ZHANG, AND J. CHANUSSOT, *Tensor de-*  
 679 *compositions for hyperspectral data processing in remote sensing: A comprehensive review*,  
 680 arXiv preprint arXiv:2205.06407, (2022).
- 681 [44] Q. WEI, J. M. BIOUCAS-DIAS, N. DOBIGEON, AND J.-Y. TOURNERET, *Multiband image fusion*  
 682 *based on spectral unmixing*, *IEEE Trans. Geosci. Remote Sens.*, 54 (2016), pp. 7236–7249.
- 683 [45] Q. WEI, N. DOBIGEON, AND J.-Y. TOURNERET, *Fast fusion of multi-band images based on*  
 684 *solving a Sylvester equation*, *IEEE Trans. Image Process.*, 24 (2015), pp. 4109–4121.
- 685 [46] N. YOKOYA, C. GROHNFELDT, AND J. CHANUSSOT, *Hyperspectral and multispectral data fusion:*  
 686 *A comparative review of the recent literature*, *IEEE Trans. Geosci. Remote Sens.*, 5 (2017),  
 687 pp. 29–56.
- 688 [47] N. YOKOYA, T. YAIRI, AND A. IWASAKI, *Coupled Nonnegative Matrix Factorization Unmixing*  
 689 *for Hyperspectral and Multispectral Data Fusion*, *IEEE Trans. Geosci. Remote Sens.*, 50  
 690 (2012), pp. 528–537, <https://doi.org/10.1109/TGRS.2011.2161320>.
- 691 [48] V. ZARZOSO, *Parameter estimation in block term decomposition for noninvasive atrial fibril-*  
 692 *lation analysis*, in 2017 IEEE 7th International Workshop on Computational Advances in  
 693 Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2017, pp. 1–5.
- 694 [49] G. ZHANG, X. FU, K. HUANG, AND J. WANG, *Hyperspectral super-resolution: A coupled non-*  
 695 *negative block-term tensor decomposition approach*, in 2019 IEEE CAMSAP, 2019. Guade-  
 696 loupe, West Indies.