



**HAL**  
open science

# Hyperspectral Super-Resolution with Coupled Tucker Approximation: Recoverability and SVD-based algorithms

Clémence Prévost, Konstantin Usevich, Pierre Comon, David Brie

► **To cite this version:**

Clémence Prévost, Konstantin Usevich, Pierre Comon, David Brie. Hyperspectral Super-Resolution with Coupled Tucker Approximation: Recoverability and SVD-based algorithms. 2019. hal-01911969v3

**HAL Id: hal-01911969**

**<https://hal.science/hal-01911969v3>**

Preprint submitted on 11 Apr 2019 (v3), last revised 10 Jan 2020 (v4)

**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.

# Hyperspectral Super-Resolution with Coupled Tucker Approximation: Recoverability and SVD-based algorithms

Clémence Prévost, *Student Member, IEEE*, Konstantin Usevich\*, *Member, IEEE*,  
Pierre Comon, *Fellow, IEEE*, and David Brie *Member, IEEE*,

**Abstract**—We propose a novel approach for hyperspectral super-resolution, that is based on low-rank tensor approximation for a coupled low-rank multilinear (Tucker) model. We show that the correct recovery holds for a wide range of multilinear ranks. For coupled tensor approximation, we propose two SVD-based algorithms that are simple and fast, but with a performance comparable to the state-of-the-art methods. The approach is applicable to the case of unknown spatial degradation and to the pansharpening problem.

**Index Terms**—hyperspectral super-resolution, low-rank tensor approximation, data fusion, recovery, identifiability

## I. INTRODUCTION

The problem of *hyperspectral super-resolution* (HSR) [1] has recently attracted much interest from the signal processing community. It consists in fusing a multispectral image (MSI), which has a good spatial resolution but few spectral bands, and a hyperspectral image (HSI), whose spatial resolution is lower than that of MSI. The aim is to recover a *super-resolution image* (SRI), which possesses both good spatial and spectral resolution. This problem is closely related to hyperspectral pansharpening [2], [3], where HSI is fused with a panchromatic image (i.e. an “MSI” with one spectral band).

Many methods were developed for the HSR problem, including coupled nonnegative matrix factorization [4] (CNMF), methods based on solving Sylvester equation [5], Bayesian approaches (HySure [6]), FUMI [7], to name a few. Motivated by the widely used linear mixing model, most of these methods are based on a coupled low-rank factorization of the matricized hyperspectral and multispectral images.

Recently, a promising tensor-based method was proposed that makes use of the inherent 3D nature of HSI [8]. Assuming that the super-resolution image itself admits a low-rank canonical polyadic (CP) decomposition (CPD), the HSR is reformulated as a coupled CP approximation. An alternating least squares (ALS) algorithm (STEREO) is proposed, achieving reconstruction performance that is competitive with the state of the art. The key property underlying the approach of

[8] is that the coupled CPD is identifiable, hence the SRI can be uniquely recovered. This approach was also successfully used for a super-resolution problem in medical imaging [9]. Still, it has several drawbacks: for instance, the appropriate rank of the CPD is not known *a priori* and may be unrelated to the number of materials; the rank can be also very large, which may affect computational complexity and convergence.

In some cases, the spatial degradation operator is unknown, therefore blind algorithms are needed. A blind version of STEREO was proposed in [8] that also uses an ALS algorithm for a coupled CP model. In [10], a simple algorithm (SCUBA) based on a single CPD of the MSI tensor and a truncated SVD of the unfolding of the HSI is introduced. A key idea proposed in [10] is to use local approximations by splitting the data cubes into separate blocks. This algorithm outperforms blind STEREO and other state-of-the-art algorithms. It also does not require separability of the spatial degradation operator.

In this paper, we propose to use another type of low-rank tensor factorization: multilinear (also known as Tucker) factorization. By assuming that the super-resolution image has approximately low multilinear rank, we reformulate the HSR problem as a coupled Tucker approximation. First, we propose two closed-form SVD-based algorithms: the first is inspired by the higher-order SVD [11] and the second (blind) is inspired by [10]. Second, we prove that, although the Tucker decomposition is not identifiable<sup>1</sup>, the SRI can be uniquely recovered for a wide range of multilinear ranks. Our experiments on a number of simulated and semi-real examples, show that the proposed algorithms have a performance comparable to those of [8] and [10], but the computational cost is much lower. Also, the proposed approach is applicable to hyperspectral pansharpening (unlike [8], which requires the MSI to have at least two spectral bands). Finally, the algorithms can accurately reconstruct spectral signatures, which is of prime importance for further processing of the HSR image.

A short version of this work [12] appears in ICASSP 2019, presenting the SCOTT algorithm and part of the simulations. The current paper additionally includes new blind algorithms, detailed analysis of the model and the algorithms, proof of the theorem for recoverability, new simulations for synthetic and semi-real data, examples on recovery of spectral signatures.

This paper is organized as follows. In Section II, we introduce our notation, define basic tensor decomposition op-

This work was partially supported by the Agence Nationale de la Recherche under grant OPTIFIN (ANR-15-CE10-0007).

C. Prévost, K. Usevich and D. Brie are with Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Boulevard des Aiguillettes, BP 70239, F-54506 Vandoeuvre-lès-Nancy, France. P. Comon is with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France. Email: clemence.prevast@univ-lorraine.fr, konstantin.usevich@univ-lorraine.fr, pierre.comon@gipsa-lab.fr, david.brie@univ-lorraine.fr. Fax: +33 383684437. Tel.: +33 372745313 (K. Usevich).

\*Corresponding author.

<sup>1</sup>This is the reason why the Tucker model was discarded in [8] as a potential model for hyperspectral super-resolution.

erations and recall the HSR problem. In Section III, we recall the CP-based model and the STEREO algorithm proposed in [8]. Section IV contains our proposed coupled Tucker model and SVD-based algorithms (SCOTT and B-SCOTT) for tensor approximation. In Section V we prove our main recoverability result for the coupled Tucker model. Section VI contains the numerical experiments.

## II. BACKGROUND AND NOTATION

### A. Basic notation

In this paper we mainly follow [13] in what concerns the tensor notation (see also [14]). The following fonts are used: lowercase ( $a$ ) or uppercase ( $I$ ) plain font for scalars, boldface lowercase ( $\mathbf{a}$ ) for vectors, uppercase boldface ( $\mathbf{A}$ ) for matrices, and calligraphic ( $\mathcal{A}$ ) for  $N$ -D arrays (tensors). Vectors are, by convention, one-column matrices. The elements of vectors/matrices/tensors are accessed as  $a_i$ ,  $A_{i,j}$  and  $\mathcal{A}_{i_1, \dots, i_N}$  respectively.  $\mathbb{R}$  stands for the real line.

For a matrix  $\mathbf{A}$ , we denote its transpose and Moore-Penrose pseudoinverse as  $\mathbf{A}^\top$  and  $\mathbf{A}^\dagger$  respectively. The notation  $\mathbf{I}_M$  is used for the  $M \times M$  identity matrix and  $\mathbf{0}_{L \times K}$  for the  $L \times K$  matrix of zeroes. We use the symbol  $\boxtimes$  for the Kronecker product of matrices (in order to distinguish it from the tensor product  $\otimes$ ), and  $\odot$  for the Khatri-Rao product.

For a matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , we denote by  $\sigma_{max}(\mathbf{X})$  and  $\sigma_{min}(\mathbf{X})$  the largest and the smallest of the  $\min(m, n)$  singular values of  $\mathbf{X}$ . We also denote by  $\text{tSVD}_R(\mathbf{X}) \in \mathbb{R}^{n \times R}$  a matrix containing  $R$  leading right singular vectors of  $\mathbf{X}$ .

We use  $\text{vec}\{\cdot\}$  for the standard column-major vectorization of a tensor or a matrix. Operator  $\bullet_p$  denotes contraction on the  $p$ th index of a tensor; when contracted with a matrix, summation is always performed on the second index of the matrix, e.g.,  $[\mathcal{A} \bullet_1 \mathbf{M}]_{ijk} = \sum_\ell \mathcal{A}_{\ell jk} M_{i\ell}$ . For a tensor  $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ , its first unfolding is denoted by  $\mathbf{Y}^{(1)} \in \mathbb{R}^{JK \times I}$ .

### B. Tensor decompositions

For a tensor  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$  and matrices  $\mathbf{U} \in \mathbb{R}^{I \times R_1}$ ,  $\mathbf{V} \in \mathbb{R}^{J \times R_2}$  and  $\mathbf{W} \in \mathbb{R}^{K \times R_3}$ , the following shorthand notation is used for the multilinear product:

$$[\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W}] = \mathcal{G} \bullet_1 \mathbf{U} \bullet_2 \mathbf{V} \bullet_3 \mathbf{W}. \quad (1)$$

which means that the  $(i, j, k)$ th entry of the above array is

$$\sum_{pqr} G_{pqr} U_{ip} V_{jq} W_{kr}.$$

If  $\mathcal{Y} = [\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W}]$ , the following identities hold for its vectorization and unfoldings, respectively:

$$\begin{aligned} \text{vec}\{\mathcal{Y}\} &= (\mathbf{W} \boxtimes \mathbf{V} \boxtimes \mathbf{U}) \text{vec}\{\mathcal{G}\}, \\ \mathbf{Y}^{(1)} &= (\mathbf{W} \boxtimes \mathbf{V}) \mathbf{G}^{(1)} \mathbf{U}^\top. \end{aligned}$$

If, in addition,

$$R_1 = \text{rank}\{\mathbf{Y}^{(1)}\}, R_2 = \text{rank}\{\mathbf{Y}^{(2)}\}, R_3 = \text{rank}\{\mathbf{Y}^{(3)}\},$$

then the multilinear product is called Tucker decomposition of  $\mathcal{Y}$  and  $(R_1, R_2, R_3)$  are called the multilinear ranks.

For matrices  $\mathbf{A} \in \mathbb{R}^{I \times F}$ ,  $\mathbf{B} \in \mathbb{R}^{J \times F}$ ,  $\mathbf{C} \in \mathbb{R}^{K \times F}$ , we will use a shorthand notation for a polyadic decomposition (sometimes also called rank-decomposition)

$$[\mathbf{A}, \mathbf{B}, \mathbf{C}] = [\mathcal{I}_F; \mathbf{A}, \mathbf{B}, \mathbf{C}]$$

where  $\mathcal{I}_F \in \mathbb{R}^{F \times F \times F}$  is a diagonal tensor of ones. In other words, if  $\mathcal{Y} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$ , then

$$\mathcal{Y}_{ijk} = \sum_r A_{ir} B_{jr} C_{kr};$$

moreover, the first unfolding can be expressed as

$$\mathbf{Y}^{(1)} = (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^\top.$$

Finally, if  $F$  is minimal,  $\mathcal{Y} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$  is called canonical polyadic (CP) decomposition and  $F$  is called the tensor rank.

### C. Hyperspectral super-resolution and degradation model

We consider a multispectral image (MSI) cube  $\mathcal{Y}_M \in \mathbb{R}^{I \times J \times K_M}$  and a hyperspectral image (HSI) cube  $\mathcal{Y}_H \in \mathbb{R}^{I_H \times J_H \times K}$  acquired from existing sensors (for instance, LANDSAT or QuickBird). The spectral resolution of MSI is lower than that of HSI ( $K_M \ll K$ ), while its spatial resolution is higher ( $I > I_H$ ,  $J > J_H$ ). The acquired MSI and HSI usually represent the same target, and  $\mathcal{Y}_M$  and  $\mathcal{Y}_H$  are viewed as two degraded versions of a single super-resolution image (SRI) data cube  $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ . The hyperspectral data fusion problem [1] consists in recovering SRI  $\mathcal{Y}$  from  $\mathcal{Y}_M$  and  $\mathcal{Y}_H$ .

In this paper, as in [8], we adopt the following degradation model, that can be compactly written as contraction of SRI:

$$\begin{cases} \mathcal{Y}_M &= \mathcal{Y} \bullet_3 \mathbf{P}_M + \mathcal{E}_M, \\ \mathcal{Y}_H &= \mathcal{Y} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \mathcal{E}_H, \end{cases} \quad (2)$$

where  $\mathcal{E}_M$ ,  $\mathcal{E}_H$  denote the noise terms,  $\mathbf{P}_M \in \mathbb{R}^{K_M \times K}$  is the spectral degradation matrix (for example, a selection-averaging matrix), and  $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$ ,  $\mathbf{P}_2 \in \mathbb{R}^{J_H \times J}$  are the spatial degradation matrices, i.e. we assume (for simplicity) that the spatial degradation is separable. For example, the commonly accepted Wald's protocol [15] uses separable Gaussian blurring and downsampling in both spatial dimensions.

## III. CP-BASED DATA FUSION

In [8] it was proposed to model the SRI data cube as a tensor with low tensor rank, i.e.  $\mathcal{Y} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$ , where  $\mathbf{A} \in \mathbb{R}^{I \times F}$ ,  $\mathbf{B} \in \mathbb{R}^{J \times F}$  and  $\mathbf{C} \in \mathbb{R}^{K \times F}$  are the factor matrices of the CPD and  $F$  is the tensor rank.

### A. The case of known spatial degradation (STEREO)

In this subsection, we consider only the case when the degradation matrices  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_M$  are known; the case of unknown degradation matrices is postponed to Section III-B. In this case, the HSR problem can be formulated as

$$\underset{\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}}{\text{minimize}} \quad f_{CP}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}), \quad (3)$$

where  $f_{CP}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}) =$

$$\|\mathcal{Y}_H - [\mathbf{P}_1 \hat{\mathbf{A}}, \mathbf{P}_2 \hat{\mathbf{B}}, \hat{\mathbf{C}}]\|_F^2 + \lambda \|\mathcal{Y}_M - [\hat{\mathbf{A}}, \hat{\mathbf{B}}, \mathbf{P}_M \hat{\mathbf{C}}]\|_F^2,$$

which is a coupled CP approximation problem. For the case when there is no noise ( $\mathcal{E}_H, \mathcal{E}_M = \mathbf{0}$ ), the coupled CP model is (generically) identifiable if

$$F \leq \min\{2^{\lfloor \log_2(K_M J) \rfloor - 2}, I_H J_H\},$$

see [8] for a proof and details of this condition.

To solve (3), an alternating optimization algorithm is proposed in [8], called STEREO (Super-resolution TENSOR RE-construction), as it is described in Algorithm 1.

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_1, P_2, P_M; F, A_0 \in \mathbb{R}^{I \times F}$ ,  
 $B_0 \in \mathbb{R}^{J \times F}, C_0 \in \mathbb{R}^{K \times F}$   
**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$   
**for**  $k = 1 : n$  **do**  
     $A_k \leftarrow \underset{A}{\operatorname{argmin}} f_{CP}(A, B_{k-1}, C_{k-1})$ ,  
     $B_k \leftarrow \underset{B}{\operatorname{argmin}} f_{CP}(A_k, B, C_{k-1})$ ,  
     $C_k \leftarrow \underset{C}{\operatorname{argmin}} f_{CP}(A_k, B_k, C)$ ,  
**end**  
 $\hat{\mathcal{Y}} \leftarrow \llbracket A_n, B_n, C_n \rrbracket$ .

**Algorithm 1: STEREO**

The updates of the factor matrices in Algorithm 1 can be computed by using efficient solvers for the (generalized) Sylvester equation [16], [17]. For example, the total cost of one iteration (updating  $A, B, C$ ) in Algorithm 1 becomes

- $O(IJK_M F + I_H J_H K F)$  flops for computing the right hand sides in the least-squares subproblems.
- $O(I^3 + J^3 + K^3 + F^3)$  flops for solving Sylvester equations;

For more details on solving Sylvester equations, see<sup>2</sup> [8, Appendix E] and Appendix A.

The initial values in Algorithm 1 are chosen as

$$\llbracket A_0, B_0, \tilde{C}_0 \rrbracket = \operatorname{CPD}_F(\mathcal{Y}_M),$$

$$Y_M^{(3)} = (P_2 B_0 \odot P_1 A_0) C_0^T, \quad \tilde{C}_0 = P_M C_0,$$

where  $\operatorname{CPD}_F$  stands for a rank- $F$  CP approximation<sup>3</sup> and the last two equations are solved in the least squares sense.

### B. The case of unknown spatial degradation

In this subsection, we recall the CP-based methods for the HSR problem in the case when the spatial degradation matrices  $P_1, P_2$  are unknown, proposed in [8] and [10]. The first solution, called Blind-STEREO was to consider the following coupled CP approximation problem:

$$\min_{\hat{A}, \hat{B}, \hat{C}} \|\mathcal{Y}_H - \llbracket \hat{A}, \hat{B}, \hat{C} \rrbracket\|_F^2 + \lambda \|\mathcal{Y}_M - \llbracket \hat{A}, \hat{B}, P_M \hat{C} \rrbracket\|_F^2,$$

where the estimated SRI is computed as  $\hat{\mathcal{Y}} = \llbracket \hat{A}, \hat{B}, \hat{C} \rrbracket$ . The conditions for correct recovery were established in [8].

In [10], an alternative approach was proposed, that uses standard CP approximation of  $\mathcal{Y}_M$  together with an SVD of  $Y_H^{(3)}$ , and a least squares problem. This approach, which does not necessary need separability of the spatial degradation operation, is summarized in Algorithm 2.

<sup>2</sup>Note that in [8, Appendix E] the cost of solving the Sylvester equation is stated as  $O(I^3)$  and not  $O(I^3 + F^3)$  as in [16].

<sup>3</sup>A low tensor rank approximation does not always exist in general, but is guaranteed to exist if all terms are imposed to be entry-wise nonnegative [18].

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_M; R, F$   
**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$   
Compute CP approximation:  $\llbracket \hat{A}, \hat{B}, \tilde{C} \rrbracket = \operatorname{CPD}_F(\mathcal{Y}_M)$   
 $Z \leftarrow \operatorname{tSVD}_R(Y_H^{(3)})$   
 $\hat{C} \leftarrow Z(P_M Z)^{\dagger} \tilde{C}$   
 $\hat{\mathcal{Y}} \leftarrow \llbracket \hat{A}, \hat{B}, \hat{C} \rrbracket$ .

**Algorithm 2: Hybrid algorithm of [10]**

As noted in [10],  $\mathcal{Y} = \llbracket A, B, C \rrbracket$  can be uniquely recovered only if  $\operatorname{rank}\{C\} = R$  does not exceed the number  $K_M$  of spectral bands in MSI. To overcome this limitation, in [10] it was proposed to apply Algorithm 2 to non-overlapping subblocks of MSI and HSI (based on the hypothesis that only a small number of materials are active in a smaller block). This is summarized in Algorithm 3, called SCUBA in [10]. It was shown in [10] that such an algorithm outperforms Blind-STEREO, and other state-of-the-art algorithms for blind HSR.

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_M; R, F$   
**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$   
Split  $\mathcal{Y}_M, \mathcal{Y}_H$  in  $L$  blocks along spatial dimensions.  
**for**  $k = 1 : L$  **do**  
    Apply Algorithm 2 to each pair of blocks in  $\mathcal{Y}_M$ ,  
     $\mathcal{Y}_H$ , and store the result in the corresponding block  
    of  $\hat{\mathcal{Y}}$ .  
**end**

**Algorithm 3: SCUBA**

## IV. TUCKER-BASED DATA FUSION

### A. Model and approximation problem

In this paper, we propose a Tucker-based coupled model<sup>4</sup> as an alternative to STEREO. Let  $\mathbf{R} = (R_1, R_2, R_3)$  be the multilinear ranks of the SRI  $\mathcal{Y}$ , and let  $\mathcal{Y} = \llbracket \mathcal{G}; U, V, W \rrbracket$  be its Tucker decomposition, where  $U \in \mathbb{R}^{I \times R_1}$ ,  $V \in \mathbb{R}^{J \times R_2}$  and  $W \in \mathbb{R}^{K \times R_3}$  are the factor matrices and  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$  is the core tensor.

With these notations, Equation (2) becomes

$$\begin{cases} \mathcal{Y}_M &= \llbracket \mathcal{G}; U, V, P_M W \rrbracket + \mathcal{E}_M, \\ \mathcal{Y}_H &= \llbracket \mathcal{G}; P_1 U, P_2 V, W \rrbracket + \mathcal{E}_H. \end{cases}$$

The HSR formulation is thus

$$\underset{\hat{\mathcal{G}}, \hat{U}, \hat{V}, \hat{W}}{\operatorname{minimize}} f_T(\hat{\mathcal{G}}, \hat{U}, \hat{V}, \hat{W}), \quad \text{where} \quad (4)$$

$$f_T(\hat{U}, \hat{V}, \hat{W}, \hat{\mathcal{G}}) = \|\mathcal{Y}_H - \llbracket \hat{\mathcal{G}}; P_1 \hat{U}, P_2 \hat{V}, \hat{W} \rrbracket\|_F^2 + \lambda \|\mathcal{Y}_M - \llbracket \hat{\mathcal{G}}; \hat{U}, \hat{V}, P_M \hat{W} \rrbracket\|_F^2. \quad (5)$$

### B. An SVD-based algorithm for known spatial degradation

A suboptimal solution to problem (4) can be found by a simple method inspired by the high-order SVD (HOSVD) [11]. This method, called SCOTT (Super-resolution based on COupled Tucker Tensor approximation), is given in Algorithm 4.

<sup>4</sup>Note that another method based on Tucker factorization and sparse approximation was proposed in [19]. However, no recoverability condition is available for that method.

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_1, P_2, P_M; (R_1, R_2, R_3)$ ,

**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$

1.  $\hat{U} \leftarrow \text{tSVD}_{R_1}(\mathbf{Y}_M^{(1)})$ ,  $\hat{V} \leftarrow \text{tSVD}_{R_2}(\mathbf{Y}_M^{(2)})$ ,  
 $\hat{W} \leftarrow \text{tSVD}_{R_3}(\mathbf{Y}_H^{(3)})$ ,

2.  $\hat{\mathcal{G}} \leftarrow \underset{\mathcal{G}}{\text{argmin}} f_T(\mathcal{G}, \hat{U}, \hat{V}, \hat{W})$

3.  $\hat{\mathcal{Y}} = \llbracket \hat{\mathcal{G}}; \hat{U}, \hat{V}, \hat{W} \rrbracket$ .

**Algorithm 4: SCOTT**

Step 2 of Algorithm 4 is the least squares problem

$$\underbrace{\begin{bmatrix} \hat{W} \boxtimes P_2 \hat{V} \boxtimes P_1 \hat{U} \\ \sqrt{\lambda} P_M \hat{W} \boxtimes \hat{V} \boxtimes \hat{U} \end{bmatrix}}_{\mathbf{X}} \text{vec}\{\hat{\mathcal{G}}\} \approx \underbrace{\begin{bmatrix} \text{vec}\{\mathcal{Y}_H\} \\ \sqrt{\lambda} \text{vec}\{\mathcal{Y}_M\} \end{bmatrix}}_z$$

that can be solved through normal equations of the form

$$(\mathbf{X}^\top \mathbf{X}) \text{vec}\{\hat{\mathcal{G}}\} = \mathbf{X}^\top z. \quad (6)$$

The matrix on the left-hand side of (6) can be written as

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} = & \mathbf{I}_{R_3} \boxtimes (\hat{V}^\top P_2^\top P_2 \hat{V}) \boxtimes (\hat{U}^\top P_1^\top P_1 \hat{U}) \\ & + \lambda (\hat{W}^\top P_M^\top P_M \hat{W}) \boxtimes \mathbf{I}_{R_1 R_2}, \end{aligned} \quad (7)$$

and the vector on the right-hand side is

$$\begin{aligned} \mathbf{X}^\top z = & \text{vec}\{\llbracket \mathcal{Y}_H; \hat{U}^\top P_1^\top, \hat{V}^\top P_2^\top, \hat{W}^\top \rrbracket\} \\ & + \sqrt{\lambda} \text{vec}\{\llbracket \mathcal{Y}_M; \hat{U}^\top, \hat{V}^\top, \hat{W}^\top P_M^\top \rrbracket\}. \end{aligned} \quad (8)$$

The normal equations can be viewed as a (generalized) Sylvester equation and (as in the case of STEREO) efficient solvers can be used (see Appendix A for more details). Thus the total cost of one iteration in SCOTT becomes

- $O(\min(R_1, R_2)IJK_M + R_3I_H J_H K)$  flops for computing the truncated SVDs and computing  $\mathbf{X}^\top z$ ;
- $O(\min(R_3^3 + (R_1 R_2)^3, R_1^3 + (R_2 R_3)^3))$  flops for solving the Sylvester equation.

It is easy to see that the computational complexity of SCOTT is comparable to that of one iteration of STEREO and is actually smaller if the multilinear ranks are small.

### C. An algorithm for unknown spatial degradation

In this subsection, we show that it is also possible to develop a blind SVD-based algorithm, in the same spirit as Algorithm 2. The algorithm does not need knowledge of  $P_1, P_2$  and is based on the HOSVD of the MSI tensor.

The total computational complexity of Algorithm 5 is

$$O(\min(R_1, R_2)IJK_M + R_3I_H J_H K) \text{ flops}$$

and is dominated by the cost of the truncated SVD, because step 3 is very cheap. However, a specific drawback of Algorithm 6, similarly to Algorithm 2, is that  $R_3$  should not exceed  $K_M$ , since the multilinear rank is employed in the HOSVD of subblocks of  $\mathcal{Y}_M$ .

Finally, similarly to SCUBA, we can use a block version of Algorithm 5, which we call B-SCOTT (which stands for ‘‘Blind SCOTT’’). There is no confusion, as Algorithm 5 is a special case of Algorithm 6 where the degraded image cubes are not split into blocks.

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_M; (R_1, R_2, R_3)$

**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$

1. Compute the  $(R_1, R_2, R_3)$  HOSVD of  $\mathcal{Y}_M$

$$\llbracket \hat{\mathcal{G}}; \hat{U}, \hat{V}, \hat{W} \rrbracket \stackrel{\text{HOSVD}}{\approx} \mathcal{Y}_M$$

2.  $Z \leftarrow \text{tSVD}_{R_3}(\mathbf{Y}_H^{(3)})$

3.  $\hat{W} \leftarrow Z(P_M Z)^\dagger \hat{W}$

4.  $\hat{\mathcal{Y}} = \llbracket \hat{\mathcal{G}}; \hat{U}, \hat{V}, \hat{W} \rrbracket$ .

**Algorithm 5: Blind version of SCOTT**

**input :**  $\mathcal{Y}_M, \mathcal{Y}_H, P_M; (R_1, R_2, R_3)$

**output:**  $\hat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$

Split  $\mathcal{Y}_M, \mathcal{Y}_H$  in  $L$  blocks along spatial dimensions.

**for**  $k = 1 : L$  **do**

    Apply Algorithm 5 to each pair of blocks in  $\mathcal{Y}_M$ ,

$\mathcal{Y}_H$ , and store the result in the corresponding block

    of  $\hat{\mathcal{Y}}$ .

**end**

**Algorithm 6: B-SCOTT (block version of Algorithm 5)**

## V. RECOVERABILITY OF THE TUCKER MODEL

In this section, we establish conditions for correct SRI tensor recovery in the coupled Tucker model. The proof of such conditions for the CP model in [8] relied on the uniqueness (identifiability) property of the CPD. We show that, although the Tucker decomposition is not unique, the correct recovery is still possible. In particular, there may exist cases where the CPD in [8] is not unique but the SRI tensor is still uniquely recovered using the CP model. Therefore, we believe that is more appropriate to use the term ‘‘recoverability of the SRI’’ rather than ‘‘identifiability of the SRI’’ used in [8].

### A. Deterministic noiseless recovery conditions

We begin with a deterministic result on recoverability.

**Theorem V.1.** *Let a Tucker decomposition of  $\mathcal{Y}$  be*

$$\mathcal{Y} = \llbracket \mathcal{G}; U, V, W \rrbracket, \quad (9)$$

where  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ , and  $U \in \mathbb{R}^{I \times R_1}$ ,  $V \in \mathbb{R}^{J \times R_2}$ ,  $W \in \mathbb{R}^{K \times R_3}$ . We also assume that  $\mathcal{E}_M, \mathcal{E}_H = \mathbf{0}$  in (2).

1) *If*

$$\begin{aligned} \text{rank}\{\mathbf{Y}_M^{(1)}\} = R_1, \text{rank}\{\mathbf{Y}_M^{(2)}\} = R_2, \\ \text{rank}\{\mathbf{Y}_H^{(3)}\} = R_3, \end{aligned} \quad (10)$$

*and one of the following conditions holds true:*

- either  $\text{rank}\{P_1 U\} = R_1$  and  $\text{rank}\{P_2 U\} = R_2$ ;*
- or  $\text{rank}\{P_M W\} = R_3$ .*

*Then there exists only one  $\hat{\mathcal{Y}}$  such that  $\hat{\mathcal{Y}} \bullet_3 P_M = \mathcal{Y}_M$  and  $\hat{\mathcal{Y}} \bullet_1 P_1 \bullet_2 P_2 = \mathcal{Y}_H$ .*

2) *If none of the conditions a) and b) is satisfied, then there exist infinitely many  $\hat{\mathcal{Y}}$  such that  $\hat{\mathcal{Y}} \bullet_3 P_M = \mathcal{Y}_M$  and  $\hat{\mathcal{Y}} \bullet_1 P_1 \bullet_2 P_2 = \mathcal{Y}_H$ ; in fact,  $\|\hat{\mathcal{Y}} - \mathcal{Y}\|$  can be arbitrary large.*

*Proof.* First of all, we note that by [20, Theorem 13.16], the singular values of the matrix  $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_{R_3} \boxtimes \mathbf{A} + \mathbf{D} \boxtimes \mathbf{I}_{R_1 R_2}$  in (7) are all sums of the pairs of eigenvalues of

$$\underbrace{(\widehat{\mathbf{V}}^\top \mathbf{P}_2^\top \mathbf{P}_2 \widehat{\mathbf{V}})}_A \boxtimes \underbrace{(\widehat{\mathbf{U}}^\top \mathbf{P}_1^\top \mathbf{P}_1 \widehat{\mathbf{U}})}_D, \lambda \widehat{\mathbf{W}}^\top \mathbf{P}_M^\top \mathbf{P}_M \widehat{\mathbf{W}}. \quad (11)$$

We also assume without loss of generality that  $\mathbf{U}, \mathbf{V}, \mathbf{W}$  have orthonormal columns.

- *Proof of 2)* Assume that  $\text{rank}\{\mathbf{P}_1 \mathbf{U}\} \text{rank}\{\mathbf{P}_2 \mathbf{U}\} < R_1 R_2$  and  $\text{rank}\{\mathbf{P}_M \mathbf{W}\} < R_3$ . If we set  $\widehat{\mathbf{U}} = \mathbf{U}$ ,  $\widehat{\mathbf{V}} = \mathbf{V}$ ,  $\widehat{\mathbf{W}} = \mathbf{W}$ , then  $\text{rank}\{\mathbf{A}\} < R_1 R_2$ ,  $\text{rank}\{\mathbf{D}\} < R_3$  and  $\text{rank}\{\mathbf{X}^\top \mathbf{X}\} < R_1 R_2 R_3$ . Therefore the system (6) is underdetermined, and there is an infinite number of solutions  $\widehat{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ . Note that if we define  $\widehat{\mathcal{Y}} = \llbracket \widehat{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$ , then it is an admissible solution, i.e.,  $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_M = \mathcal{Y}_M$  and  $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ . On the other hand, due to orthogonality of the bases,  $\|\widehat{\mathcal{Y}} - \mathcal{Y}\|_F = \|\widehat{\mathcal{G}} - \mathcal{G}\|_F$ , which can be made arbitrary large due to nonuniqueness of the solution to (6).
- *Proof of 1)* Let us choose  $\widehat{\mathbf{U}} \in \mathbb{R}^{I \times R_1}$ ,  $\widehat{\mathbf{V}} \in \mathbb{R}^{J \times R_2}$ , and  $\widehat{\mathbf{W}} \in \mathbb{R}^{K \times R_3}$  to be orthogonal bases of the row spaces of  $\mathbf{Y}_M^{(1)}$ ,  $\mathbf{Y}_M^{(2)}$  and  $\mathbf{Y}_H^{(3)}$  respectively. First, by (10), the rank of unfoldings does not drop after degradation, hence

$$\widehat{\mathbf{U}} = \mathbf{U} \mathbf{Q}_U, \widehat{\mathbf{V}} = \mathbf{V} \mathbf{Q}_V, \widehat{\mathbf{W}} = \mathbf{W} \mathbf{Q}_W,$$

where  $\mathbf{Q}_U, \mathbf{Q}_V, \mathbf{Q}_W$  are some rotation matrices. Next, due to conditions on the ranks of  $\mathbf{P}_1 \mathbf{U}$ ,  $\mathbf{P}_2 \mathbf{U}$  and  $\mathbf{P}_M \mathbf{W}$ , we get that  $\text{rank}\{\mathbf{X}^\top \mathbf{X}\} = R_1 R_2 R_3$  because of (11). Hence the solution  $\widehat{\mathcal{G}}$  of (6) is unique. Finally, we note that the reconstructed tensor can be expressed as

$$\text{vec}\{\widehat{\mathcal{Y}}\} = (\widehat{\mathbf{W}} \boxtimes \widehat{\mathbf{V}} \boxtimes \widehat{\mathbf{U}}) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z},$$

where the right-hand side does not depend on the rotation matrices  $\mathbf{Q}_U, \mathbf{Q}_V$ , and  $\mathbf{Q}_W$  due to the definition of  $\mathbf{X}$ . Hence, the reconstructed tensor  $\widehat{\mathcal{Y}}$  is unique.  $\square$

**Corollary V.2.** *If the conditions of Theorem V.1 (part 1) hold, then any minimizer of (4) recovers  $\mathcal{Y}$ , i.e.*

$$\mathcal{Y} = \llbracket \widehat{\mathcal{G}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}} \rrbracket.$$

*In addition, Algorithm 4 recovers  $\mathcal{Y}$  for all cases of recoverability in Theorem V.1, and Algorithm 5 if  $R_3 \leq K_M$ .*

Theorem V.1, apart from results for the Tucker model, shows that the coupled CPD can also recover the SRI tensor uniquely, even for the cases when the CPD itself is not unique.

**Corollary V.3** (Recoverability with partial uniqueness). *Assume that the SRI has a CPD  $\mathcal{Y} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$  of rank  $F \leq \min(I_H, J_H)$ , such that*

$$\text{rank}\{\mathbf{A}\} = \text{rank}\{\mathbf{P}_1 \mathbf{A}\} = \text{rank}\{\mathbf{B}\} = \text{rank}\{\mathbf{P}_2 \mathbf{B}\} = F$$

*and  $\mathbf{P}_M \mathbf{C}$  does not have zero columns. We also assume that  $\mathcal{E}_M, \mathcal{E}_H = \mathbf{0}$  in (2). Then any minimizer of (3) recovers  $\mathcal{Y}$ .*

*Proof.* Since the original factors  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  yield zero error in (3), hence any global minimizer  $(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}})$  of (3), satisfies

$$\llbracket \mathbf{P}_1 \widehat{\mathbf{A}}, \mathbf{P}_2 \widehat{\mathbf{B}}, \widehat{\mathbf{C}} \rrbracket = \mathcal{Y}_H \text{ and } \llbracket \widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \mathbf{P}_M \widehat{\mathbf{C}} \rrbracket = \mathcal{Y}_M.$$

Due to the conditions of the corollary,  $\llbracket \mathbf{P}_1 \mathbf{A}, \mathbf{P}_2 \mathbf{B}, \mathbf{C} \rrbracket$  and  $\llbracket \mathbf{A}, \mathbf{B}, \mathbf{P}_M \mathbf{C} \rrbracket$  satisfy partial uniqueness conditions in [21, Theorem 2.2]. Hence (after permutations and rescaling of factors), we have  $\widehat{\mathbf{C}} = \mathbf{C}$  and

$$\text{rank}\{\mathbf{C}\} = \text{rank}\{\widehat{\mathbf{C}}\} = \text{rank}\{\mathbf{Y}_H^{(3)}\} = R_3.$$

Moreover, since

$$\mathbf{Y}_M^{(1)} = (\mathbf{P}_M \mathbf{C} \odot \mathbf{B}) \mathbf{A}^\top, \quad \mathbf{Y}_M^{(2)} = (\mathbf{P}_M \mathbf{C} \odot \mathbf{A}) \mathbf{B}^\top,$$

and  $\mathbf{P}_M \mathbf{C}$  does not have zero columns, we have that

$$\begin{aligned} \text{rank}\{\mathbf{A}\} &= \text{rank}\{\widehat{\mathbf{A}}\} = \text{rank}\{\mathbf{Y}_M^{(1)}\} = R_1 = F, \\ \text{rank}\{\mathbf{B}\} &= \text{rank}\{\widehat{\mathbf{B}}\} = \text{rank}\{\mathbf{Y}_M^{(2)}\} = R_2 = F. \end{aligned}$$

Therefore, both  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$  and  $(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}})$  are particular solutions of Problem (4) with an additional constraint that the tensor rank of  $\mathcal{G}$  is at most  $F$ . Since, by Theorem V.1, any solution of (4) recovers  $\mathcal{Y}$  uniquely, the proof is complete.  $\square$

Note that the conditions of Corollary V.3 most likely can be relaxed, for example, by using Kruskal ranks and more general formulation in [21, Theorem 2.1] (see also [22]).

### B. Noiseless ecoverability for generic tensors

From the deterministic recovery conditions, we can establish the generic recoverability results.

**Theorem V.4.** *Assume that  $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$ ,  $\mathbf{P}_2 \in \mathbb{R}^{J_H \times J}$ , and  $\mathbf{P}_M \in \mathbb{R}^{K_M \times K}$  are fixed full row-rank matrices. Let  $\mathcal{Y}$  have decomposition (9), where  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ ,  $R_1 \leq I$ ,  $R_2 \leq J$ ,  $R_3 \leq K$ , and  $\mathbf{U}, \mathbf{V}, \mathbf{W}$  are random tensors and matrices, distributed according to an absolutely continuous probability distribution. We also assume that  $\mathcal{E}_M, \mathcal{E}_H = \mathbf{0}$  in (2).*

- 1) *If  $R_3 \leq K_M$  or  $(R_1, R_2) \leq (I_H, J_H)$  and*

$$\begin{cases} R_1 \leq \min(R_3, K_M) R_2, \\ R_2 \leq \min(R_3, K_M) R_1, \\ R_3 \leq \min(R_1, I_H) \min(R_2, J_H), \end{cases} \quad (12)$$

*then with probability 1 there exists a unique tensor  $\widehat{\mathcal{Y}}$  such that  $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_M = \mathcal{Y}_M$  and  $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ .*

- 2) *If  $R_3 > K_M$  and  $(R_1 > I_H$  or  $R_2 > J_H)$ , then the reconstruction is non-unique, i.e. there exist infinitely many  $\widehat{\mathcal{Y}}$  such that  $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_M = \mathcal{Y}_M$  and  $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ ; in fact,  $\|\widehat{\mathcal{Y}} - \mathcal{Y}\|$  can be arbitrary large.*

*Proof.* • *Proof of 2)* follows from Theorem V.1 (part 2)

- *Proof of 1)* First, without loss of generality, we can replace  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_M$  with the following of same size:

$$\tilde{\mathbf{P}}_1 = \begin{bmatrix} \mathbf{I}_{I_H} \\ \mathbf{0} \end{bmatrix}^\top, \tilde{\mathbf{P}}_2 = \begin{bmatrix} \mathbf{I}_{J_H} \\ \mathbf{0} \end{bmatrix}^\top, \tilde{\mathbf{P}}_M = \begin{bmatrix} \mathbf{I}_{K_M} \\ \mathbf{0} \end{bmatrix}^\top. \quad (13)$$

Indeed, let us explain why it is so, for example for  $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$ . There exists a nonsingular matrix<sup>5</sup>  $\mathbf{T}$  such that

$$\mathbf{P}_1 \mathbf{T} = [\mathbf{I}_{I_H} \quad \mathbf{0}].$$

If we take  $\tilde{\mathbf{U}} = \mathbf{T}^{-1} \mathbf{U}$  then  $\mathbf{P}_1 \mathbf{U} = \tilde{\mathbf{P}}_1 \tilde{\mathbf{U}}$ . Note that a nonsingular transformation preserves absolute continuity

<sup>5</sup>For example,  $\mathbf{T} = [\mathbf{P}_1^\dagger \mathbf{F}]$ , where  $\mathbf{F} \in \mathbb{R}^{I \times (I - I_H)}$ ,  $\mathbf{P}_1 \mathbf{F} = \mathbf{0}$ .

of the distribution; hence  $\mathbf{U}$  has an absolutely continuous distribution if and only if  $\tilde{\mathbf{U}}$  has one.

Therefore, under the assumptions on distribution of  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{W}$  the following implications hold with probability 1

$$\begin{aligned} R_1 \leq I_H &\Rightarrow \text{rank}\{\mathbf{U}_{1:I_H,:}\} = R_1, \\ R_2 \leq J_H &\Rightarrow \text{rank}\{\mathbf{V}_{1:J_H,:}\} = R_2, \\ R_3 \leq K_M &\Rightarrow \text{rank}\{\mathbf{W}_{1:K_M,:}\} = R_3. \end{aligned}$$

Next, we are going to show how the other set of conditions imply (10). We will prove it only for the first condition (the others are analogous). Note that the first unfolding can be written as

$$\mathbf{Y}_M^{(1)} = (\mathbf{W}_{1:K_M,:} \boxtimes \mathbf{V}) \mathbf{G}^{(1)} \mathbf{U}^T.$$

Due to the dimensions of the terms in the product, this matrix is at most rank  $R_1$ . Due to semicontinuity of the rank function,  $\mathbf{Y}_M^{(1)}$  will be generically of rank  $R_1$  if we can provide just a single example of  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{W}$ ,  $\mathbf{G}$ , achieving the condition  $\text{rank}\{\mathbf{Y}_M^{(1)}\} = R_1$ . Indeed, if  $R_1 \leq \min(R_3, K_M)R_2$ , such an example is given by

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_{R_1} \\ \mathbf{0} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \mathbf{I}_{R_2} \\ \mathbf{0} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \mathbf{I}_{R_3} \\ \mathbf{0} \end{bmatrix}, \mathbf{G}^{(1)} = \begin{bmatrix} \mathbf{I}_{R_1} \\ \mathbf{0} \end{bmatrix},$$

which completes the proof.  $\square$

We illustrate the statement of Theorem V.4 for the case  $I = J$ ,  $I_H = J_H$  and  $R_1 = R_2$ . In Figure 1 we show that the space of parameters  $(R_1, R_3)$  is split into two regions: recoverable and non-recoverable. The hatched area corresponds to the parameters where condition (12) is not satisfied.

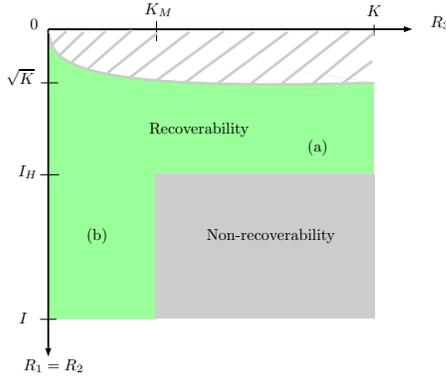


Fig. 1. Identifiability region depending on  $R_1$  and  $R_3$

**Remark V.5.** In the proof of Theorem V.4 it was shown that we can assume that the degradation operators are given in (13). In that case, the degraded tensors  $\mathcal{Y}_M$  and  $\mathcal{Y}_H$  are just the subtensors (slabs) i.e.

$$\mathcal{Y}_M = \mathcal{Y}_{:, :, 1:K_M}, \quad \mathcal{Y}_H = \mathcal{Y}_{1:I_H, 1:J_H, :}$$

Hence the recoverability of Tucker super-resolution model is equivalent to uniqueness of tensor completion, that is the recovery of  $\mathcal{Y}$  from known subtensors  $\mathcal{Y}_{:, :, 1:K_M}$  and  $\mathcal{Y}_{1:I_H, 1:J_H, :}$ , shown in Figure 2.

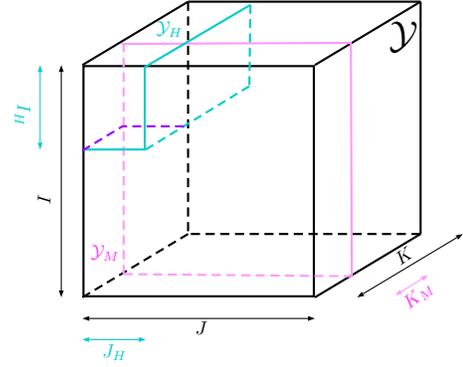


Fig. 2. Recovery of  $\mathcal{Y}$  from  $\mathcal{Y}_{:, :, 1:K_M}$  (pink),  $\mathcal{Y}_{1:I_H, 1:J_H, :}$  (blue)

## VI. EXPERIMENTAL SETUP

All simulations were run on a MacBook Pro with 2.3 GHz Intel Core i5 and 16GB RAM. The code is implemented in MATLAB. For basic tensor operations we used TensorLab 3.0 [23]. The results are reproducible and the codes are available online at [https://github.com/cprevost4/HSR\\_Tucker](https://github.com/cprevost4/HSR_Tucker).

### A. Degradation model

Experiments are conducted on a set of semi-real and synthetic examples, in which the groundtruth SRI is artificially degraded to  $\mathcal{Y}_H$  and  $\mathcal{Y}_M$  by the degradation matrices  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  and  $\mathbf{P}_M$  according to model (2).

For spatial degradation, we follow the commonly used Wald's protocol [15]. The matrices  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  are computed with a separable Gaussian blurring kernel of size  $q = 9$ . Then, downsampling is performed along each spatial dimension with a ratio  $d = 4$  between  $I, J$  and  $I_H, J_H$ , as in [8]. We refer to Appendix B for more details on the construction of  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ .

In this paper, we consider two spectral responses used to generate the spectral degradation matrix  $\mathbf{P}_M$ . In all the semi-real examples, available online at [24], the bands corresponding to water absorption are first removed as in [8]. The LANDSAT sensor spans the spectrum from 400nm to 2500nm for the HSI and produces a 6-band MSI corresponding to wavelengths 450–520nm (blue), 520–600nm (green), 630–690nm (red), 760–900nm (near-IR), 1550–1750nm (shortwave-IR) and 2050–2350nm (shortwave-IR2). The second response corresponds to a QuickBird sensor, which spans the spectrum from 430nm to 860nm for the HSI and produces a 4-band MSI which bands correspond to wavelengths 430–545nm (blue), 466–620nm (green), 590–710nm (red) and 715–918nm (near-IR). The spectral degradation matrix  $\mathbf{P}_M$  is a selection-averaging matrix that selects the common spectral bands of the SRI and MSI.

### B. Metrics

As for the experimental setup, we follow [8]; we compare the groundtruth SRI with the recovered SRI obtained by the algorithms. The main performance metric used in comparisons is reconstruction Signal-to-Noise ratio (R-SNR) used in [3]:

$$\text{R-SNR} = 10 \log_{10} \left( \frac{\|\mathcal{Y}\|_F^2}{\|\hat{\mathcal{Y}} - \mathcal{Y}\|_F^2} \right). \quad (14)$$

In addition to R-SNR, we consider different metrics from [3] described below:

$$\text{CC} = \frac{1}{IJK} \left( \sum_{k=1}^K \rho(\mathbf{y}_{:,k}, \hat{\mathbf{y}}_{:,k}) \right), \quad (15)$$

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

$$\text{SAM} = \frac{180}{\pi} \frac{1}{IJ} \sum_{n=1}^{IJ} \arccos \left( \frac{\mathbf{Y}_{n,:}^{(3)\top} \hat{\mathbf{Y}}_{n,:}^{(3)}}{\|\mathbf{Y}_{n,:}^{(3)}\|_2 \|\hat{\mathbf{Y}}_{n,:}^{(3)}\|_2} \right), \quad (16)$$

which computes the angle between original and estimated fibers;

$$\text{ERGAS} = \frac{100}{d} \sqrt{\frac{1}{IJK} \sum_{k=1}^K \frac{\|\hat{\mathbf{y}}_{:,k} - \mathbf{y}_{:,k}\|_F^2}{\mu_k^2}}, \quad (17)$$

where  $\mu_k^2$  is the mean value of  $\hat{\mathbf{y}}_{:,k}$ . 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 show the computational time for each algorithm, given by the `tic` and `toc` functions of MATLAB.

### C. Semi-real data: comparison with other methods

In this subsection, we showcase the capabilities of SCOTT and BSCOTT and compare them with state-of-the-art methods.

1) *Non-blind algorithms:* We compare the performance of non-blind algorithms (*i.e.* STEREO and SCOTT). We test various ranks for both algorithms. For STEREO, we use the implementation of [8], available online at [25]. In other subsections, we use our implementation with fast solvers for the Sylvester equations (see Appendix A). For HySure [6], the groundtruth number of materials is chosen as the number of endmembers as in [8]. This algorithm is applied in a non-blind fashion, meaning that the spatial<sup>6</sup> and spectral degradation operators are not estimated but obtained from  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  and  $\mathbf{P}_M$ . As a comparison, we also show the performance of BSCOTT when no splitting is performed.

The first dataset we consider is Indian Pines, where  $\mathbf{Y} \in \mathbb{R}^{144 \times 144 \times 200}$  is degraded by a LANDSAT sensor for the MSI and a downsampling ratio  $d = 4$  for the HSI. The results are presented in Table I and Figure 3.

Algorithm	R-SNR	CC	SAM	ERGAS	time
STEREO 50	26.91	0.89	2.26	1.03	0.85
STEREO 100	28.53	0.92	2.02	0.87	2.04
SCOTT [40,40,6]	26.39	0.89	2.32	1.06	0.17
SCOTT [30,30,16]	25.15	0.87	2.5	1.18	0.36
SCOTT [70,70,6]	27.62	0.9	2.19	0.95	0.36
SCOTT [24,24,25]	25.08	0.88	2.44	1.17	0.16
B-SCOTT [40,40,6]	25.41	0.88	2.66	1.2	0.07
HySure E = 16	28.31	0.92	2.08	0.87	15.27

TABLE I  
INDIAN PINES (NON-BLIND ALGORITHMS)

We can see that for multilinear ranks chosen in the recoverability region (see Figure 1), SCOTT yields similar

<sup>6</sup>In fact, HySure has a different, convolutional degradation model, that is not necessarily separable.

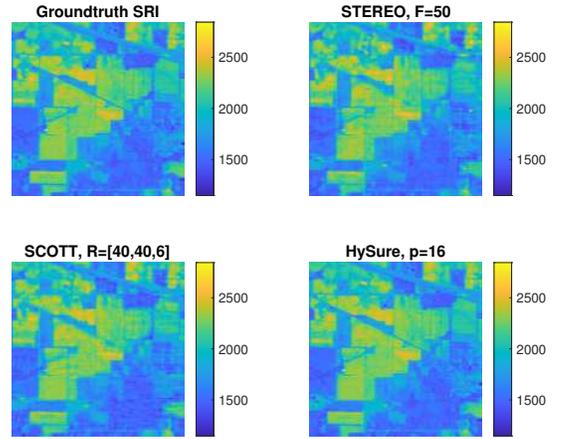


Fig. 3. Spectral slice 120 of the SRI, Indian Pines

performance to the one of STEREO with lower computation time. Moreover, contrary to [8] (where  $F = 50$  is taken for STEREO), we found out that the tensor rank  $F = 100$  yields better performance.

The other dataset is the Salinas-A scene, where  $\mathbf{Y} \in \mathbb{R}^{80 \times 84 \times 204}$  is degraded with QuickBird specifications and  $d = 4$  for the HSI. The results are presented in Table II and Figure 4.

Algorithm	R-SNR	CC	SAM	ERGAS	time
STEREO F = 50	33.4	0.97	0.91	3.35	0.92
STEREO F = 100	32.91	0.94	0.58	5.41	0.9
SCOTT [40,40,6]	31.52	0.95	0.71	4.92	0.1
SCOTT [50,50,6]	32.31	0.95	0.59	4.89	0.13
SCOTT [70,70,6]	32.89	0.95	0.48	4.89	0.28
B-SCOTT [40,40,6]	31.3	0.95	0.74	4.96	0.03
HySure E = 6	31.59	0.95	0.65	4.96	4.72

TABLE II  
SALINAS A-SCENE (NON-BLIND ALGORITHMS)

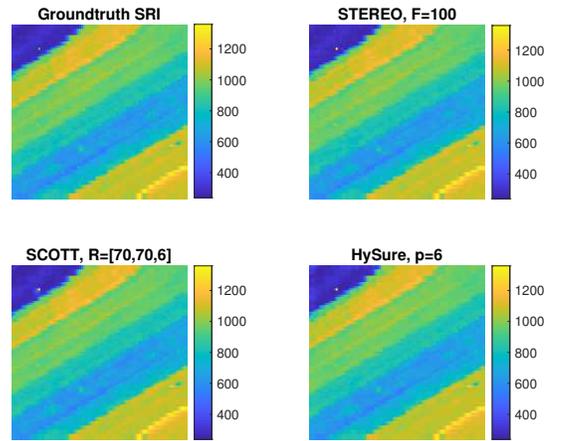


Fig. 4. Spectral slice 120 of the SRI, Salinas-A scene

In [8], CP-rank  $F = 100$  is used for STEREO. However, we found out that for STEREO, tensor rank  $F = 50$  yields better reconstruction. In Figure 4, we can see that STEREO and SCOTT can recover accurately the SRI.

2) *Blind algorithms:* We now consider the case where the spatial degradation matrices  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  are unknown and compare the performance of BSCOTT with Blind STEREO

[8], SCUBA [10] and HySure. We consider two other datasets; the first one is a portion of the Pavia University, where  $\mathcal{Y} \in \mathbb{R}^{608 \times 366 \times 103}$  is degraded with QuickBird specifications for the MSI and  $d = 4$  for the HSI. We demonstrate the results in Table III and Figure 5 for visual reconstruction. For BSCOTT, in the case where  $R = [152, 84, 3]$ , no compression is performed.

Algorithm	R-SNR	CC	SAM	ERGAS	time
SCUBA (4,4) [120,3]	25.67	0.99	3.25	1.97	17.6
B-SCOTT (4,4) [60,60,3]	23.41	0.99	3.83	2.37	0.36
B-SCOTT (4,4) [152,84,3]	26.42	0.99	2.97	1.83	0.55
B-SCOTT (4,4) [120,60,4]	25.63	0.99	3.01	1.81	0.48
SCUBA (8,8) [120,3]	26.49	0.99	2.93	1.81	49.35
B-SCOTT (8,8) [70,40,3]	26.43	0.99	2.95	1.83	0.73
B-STEREO F = 300	23.14	0.98	3.96	2.45	91.35
HySure E = 9	26.05	0.99	2.9	1.79	101.54

TABLE III

PAVIA UNIVERSITY (BLIND ALGORITHMS)

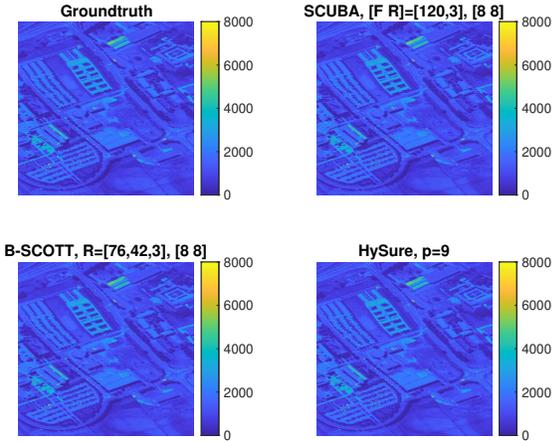


Fig. 5. Spectral slice 44 of the SRI, Pavia University

In the second case, we consider the Cuprite dataset, where  $\mathcal{Y} \in \mathbb{R}^{512 \times 614 \times 224}$  is degraded with LANDSAT specifications and  $d = 4$ . The results are presented in Table IV and Figure 6.

Algorithm	R-SNR	CC	SAM	ERGAS	time
SCUBA (4,4) [45,3]	31.7	0.97	1.12	6.57	12.1
B-SCOTT (4,4) [45,45,3]	31.91	0.97	1.08	6.57	0.76
B-SCOTT (4,4) [60,60,3]	33.02	0.98	1.03	6.58	1
SCUBA (8,8) [45,3]	34.66	0.99	0.92	6.17	34.22
B-SCOTT (8,8) [45,45,3]	34.7	0.99	0.91	6.19	1.19
B-STEREO F = 150	29.97	0.97	1.35	7.36	59.92
HySure E = 10	34.61	0.99	0.94	6.66	169.96

TABLE IV

CUPRITE (BLIND ALGORITHMS)

These two previous examples show that, for different splittings, and ranks taken from [10], BSCOTT yields the best performance. For certain multilinear ranks, it even outperforms SCUBA with lower computation time. Moreover, it outperforms Blind STEREO. In terms of visual reconstruction, our algorithm can recover accurately the details of the groundtruth SRI, even though the spatial degradation matrices are unknown.

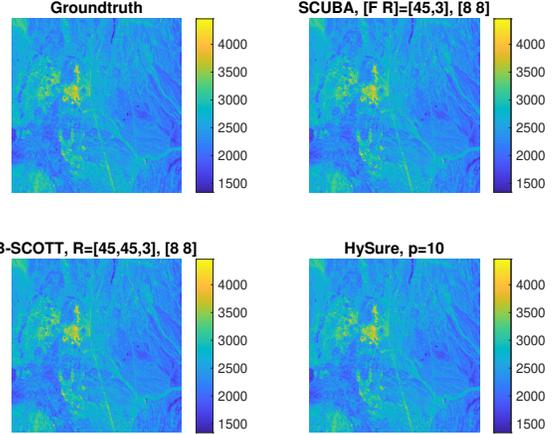


Fig. 6. Spectral slice 44 of the SRI, Cuprite

3) *Hyperspectral pansharpening*: Next, we address the pansharpening problem, which consists in fusion of a hyperspectral image and a panchromatic image (PAN)  $\mathcal{Y}_P \in \mathbb{R}^{I \times J \times 1}$ . In this case, the spectral degradation matrix is obtained by averaging over the full spectral range of the groundtruth SRI, so that  $P_M \in \mathbb{R}^{1 \times K}$ . CP-based algorithms are not applicable, since their initialization is based on the CPD of the MSI (which is a matrix in the case of PAN images). In Table V, the metrics are shown for different multilinear ranks for the Indian Pines dataset. We also compare our results to those of HySure. We can see that even though the only possible value of  $R_3$  is 1 for BSCOTT, the algorithm still manages to yield a good recovery of the SRI. On the other hand, SCOTT can also recover the SRI accurately, but is more sensitive to the choice of the multilinear rank.

Algorithm	R-SNR	CC	SAM	ERGAS	time
SCOTT [24,24,25]	20.47	0.77	4.41	1.95	0.22
SCOTT [30,30,16]	18.04	0.69	5.68	2.59	0.43
SCOTT [35,35,6]	14.61	0.54	7.84	3.89	0.89
HySure E = 16	20.67	0.76	4.24	1.99	14.39
BSCOTT (4,4) [24,24,1]	19.78	0.72	5.07	2.19	0.28
BSCOTT (4,4) [35,35,1]	19.79	0.72	5.07	2.19	0.12

TABLE V

INDIAN PINES (PANSARPENING)

#### D. CP vs. Tucker for synthetic data

In this subsection, we address the question of model identifiability versus correct recovery of the SRI. In particular, we would like to see whether identifiability of the CP model is a necessary condition to the reconstruction of  $\mathcal{Y}$ . Thus, we showcase some synthetic examples where the CP model is not identifiable, but STEREO still recovers the correct SRI. We also present cases where the CP approach fails, while the Tucker-based methods still perform correct estimation of the SRI for a wide range of multilinear ranks under recoverability conditions.

1) *Generating synthetic SRI*: First, we explain how the synthetic SRI  $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$  are generated. We consider  $N$  spectral signatures  $s_1, \dots, s_N$  obtained from the Indian Pines groundtruth data [24]. The SRI is split into  $M^2$  equal blocks along the spatial dimensions. In each  $\frac{I}{M} \times \frac{J}{M}$  block, at most one material is active, indicated by a number in

the corresponding cell of a parcel map (see Table VI for an example).

1	2
2	

TABLE VI  
PARCEL MAP FOR  $N = 2$

Formally, the SRI is computed as

$$\mathcal{Y} = \sum_{n=1}^N \mathbf{A}_n \otimes \mathbf{s}_n, \quad (18)$$

where the abundance map  $\mathbf{A}_n$  is a block matrix with Gaussians of fixed size present on the blocks corresponding to material  $n$  in the parcel map.

For instance, we consider the case presented in Table VI; the two abundance maps are

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{H} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & \mathbf{H} \\ \mathbf{H} & 0 \end{bmatrix},$$

where  $\mathbf{H}$  is a  $60 \times 60$  Gaussian with  $\sigma = 20$ . To illustrate this example, we show in Figure 7 two spectral bands of  $\mathcal{Y}$ .

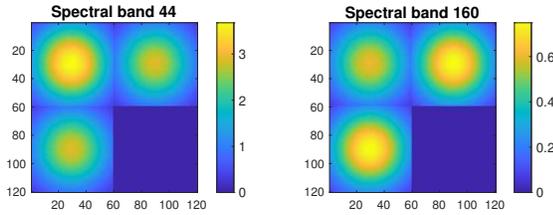


Fig. 7. Spectral bands of the synthetic SRI with  $N = 2$

Due to separability of the Gaussians,  $\mathcal{G}$  has the following multilinear decomposition with multilinear rank  $(2, 2, 2)$ .

$$\mathcal{Y} = \llbracket \mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{S} \rrbracket,$$

where  $\mathcal{G}_{:, :, 1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ,  $\mathcal{G}_{:, :, 2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ,

$$\mathbf{U} = \mathbf{V} = \begin{bmatrix} \mathbf{H} & 0 \\ 0 & \mathbf{H} \end{bmatrix} \text{ and } \mathbf{S} = [s_1 \quad s_2].$$

The tensor rank of  $\mathcal{Y}$  is equal to the tensor rank of  $\mathcal{G}$ , which is known to be equal to  $F = 3$  [26, Ex. 2], [27, Ex. 6.6], and is a well-known case where the rank-2 CP approximation does not exist in the real field [28], [29]. Thus we can expect problems with the CP-based data fusion and rank  $F = 2$ .

2) *Structured examples*: In this subsection, we present two particular examples where the CPD is not unique and we can expect computational problems due to the non-existence of low-rank CP approximation.

The first synthetic dataset is made of  $N = 2$  materials with spatial degradation ratio of  $d = 4$  for the HSI and LANDSAT specifications for the MSI (see Appendix B). In this case, introduced in the previous subsection, the tensor rank of  $\mathcal{Y}$  is at most  $F = 3$  while the multilinear rank is  $\mathbf{R} = (2, 2, 2)$ . We run STEREO for  $F$  in  $[1 : 40]$  and SCOTT for  $R_1 = R_2$  in  $[1 : I_H + 10]$  and  $R_3$  in  $[1 : 10]$  under recoverability conditions. For each algorithm, we compute the R-SNR as a function of the rank; the results are provided on Figure 8.

As a comparison, on the same plot as STEREO, we plot the results of SCOTT for  $R_3 = N$  and  $R_1 = R_2 = F$ .

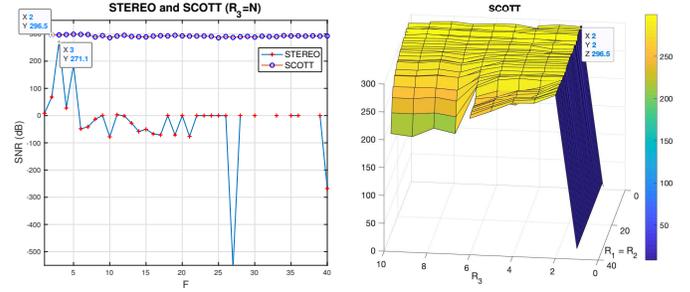


Fig. 8. R-SNR as a function of the rank

For STEREO, only rank  $F = 3$  allows for an accurate reconstruction of the SRI. For other tensor ranks, either the algorithm breaks (when no point is plotted, e.g.  $F = 32$ ) or leads to inaccurate recovery. For SCOTT, the best reconstruction error (given by R-SNR) is obtained for  $R_3 = N$  or  $R_1 = R_2 = N$ .

We also consider a bigger dataset made of  $N = 6$  materials, generated similarly to the previous example, as illustrated in Table VII and Figure 9.

1	2	3	4	5	6
2	3		5	6	
3			6		
4	5	6			
5	6				
6					

TABLE VII  
PARCEL MAP FOR  $N = 6$

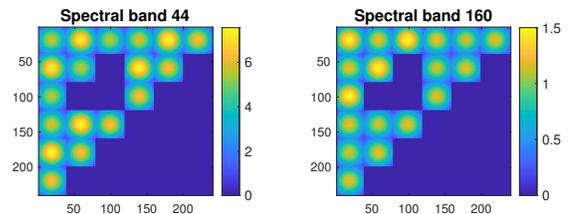


Fig. 9. Spectral bands of the synthetic SRI with  $N = 6$

In this example,  $\mathcal{Y} \in \mathbb{R}^{240 \times 240 \times 200}$  is degraded with QuickBird specifications for the MSI and  $d = 4$  for the HSI. White Gaussian noise is added to the degraded images with an input SNR of 35dB. The multilinear rank of the SRI is  $\mathbf{R} = (6, 6, 6)$ . Similarly, we run both algorithms with the same setup as in the previous example, including a comparison of STEREO and SCOTT for  $R_3 = N$ . Results are presented in Figure 10.

Due to the presence of noise, STEREO can perform data fusion accurately for a variety of tensor ranks, even though the model is not identifiable. The best R-SNR for SCOTT is obtained for  $\mathbf{R} = (6, 6, 6)$ , which is the theoretical multilinear rank of the tensor. Moreover, the best reconstruction error is obtained for  $R_3 = N$ : in this case, the performance of SCOTT is slightly better than the one of STEREO. Other values of  $R_1 = R_2$  can be used as well, according to Figure 10.



In the semi-real cases, a clear corner in the singular value curves cannot be found, because these examples do not correspond to a “low-rank signal+noise” scenario, contrary to the case of synthetic data. Moreover, the HSI and MSI are not necessarily low-rank: hence, the Tucker approach only performs a *low-rank approximation* of the data. Hence, the SVD of the unfolding does not provide as much information as for the synthetic case, in which the groundtruth data are explicitly designed to be low-rank.

2) *Influence on the reconstruction error*: Next, we consider the R-SNR and cost function  $f_T$  as functions of the multilinear rank. We run SCOTT for the ranks  $R_1 = R_2$  in  $[10 : 50]$  and  $R_3$  in  $[2 : 25]$  for which the recoverability condition holds (see Section V), and two semi-real datasets: Indian Pines and Salinas-A scene. The results are shown in Figures 15 and 16, respectively.

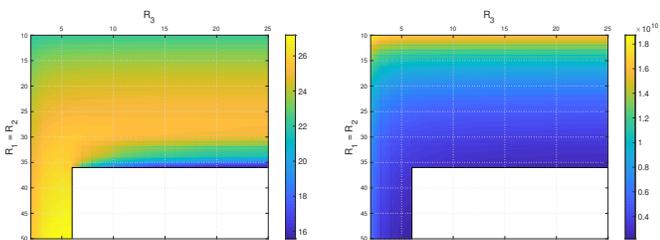


Fig. 15. R-SNR (left) and  $f_T$  (right) as functions of  $R_1$  and  $R_3$ , Indian Pines

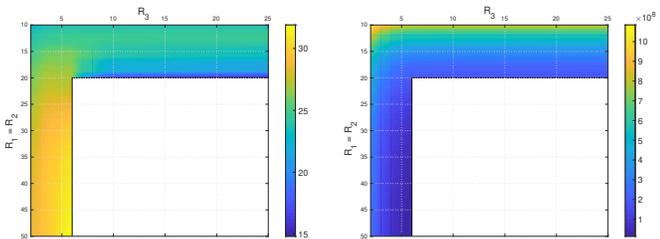


Fig. 16. R-SNR (left) and  $f_T$  (right) as functions of  $R_1$  and  $R_3$ , Salinas-A

While the cost function decreases as  $R_1$  and  $R_3$  increase, the best reconstruction error (given by R-SNR) is achieved in one of the two recoverability subregions in Fig. 1: (a) ( $R_3 \geq K_M$  and  $R_1 \leq I_H$ ) and (b) ( $R_3 \leq K_M$  and  $R_1 \geq I_H$ ). For subregion (b), the best performance is achieved when  $R_3 = K_M$  and  $R_1$  as large as possible, while for subregion (a), we notice a sharp drop of the R-SNR near  $R_1 = I_H$ .

The drop of the performance in subregion (a) can be explained by looking at the condition number of the matrix  $\mathbf{X}^T \mathbf{X}$  that is used to compute the core tensor  $\hat{\mathcal{G}}$ . For the subregion (a), due to properties of Kronecker products [20, Theorems 13.12 and 13.16], we have that

$$\begin{aligned} \text{cond}\{\mathbf{X}^T \mathbf{X}\} &:= \frac{\sigma_{\max}(\mathbf{X}^T \mathbf{X})}{\sigma_{\min}(\mathbf{X}^T \mathbf{X})} \\ &= \frac{\lambda \sigma_{\max}^2(\mathbf{P}_M \hat{\mathbf{W}}) + \sigma_{\max}^2(\mathbf{P}_1 \hat{\mathbf{U}}) \sigma_{\max}^2(\mathbf{P}_2 \hat{\mathbf{V}})}{\sigma_{\min}^2(\mathbf{P}_1 \hat{\mathbf{U}}) \sigma_{\min}^2(\mathbf{P}_2 \hat{\mathbf{V}})}. \end{aligned}$$

Note that  $\sigma_{\max}(\mathbf{P}_M \hat{\mathbf{W}})$  does not decrease when we increase  $R_3$  and  $R_3 \leq K_M$ . Hence we can get a lower bound on  $\text{cond}\{\mathbf{X}^T \mathbf{X}\}$  by setting  $R_3 = K_M$ .

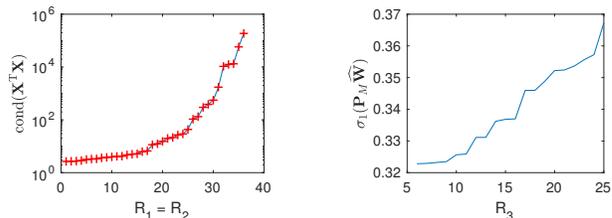


Fig. 17. Left:  $\log(\text{cond}\{\mathbf{X}^T \mathbf{X}\})$  depending on  $R_1 = R_2$  for  $R_3 = K_M$ ; right:  $\sigma_1(\mathbf{P}_M \hat{\mathbf{W}})$ , Indian Pines

In Figure 17, for the Indian Pines dataset we plot on a semi-log scale the lower bound  $\text{cond}\{\mathbf{X}^T \mathbf{X}\}$  as functions of  $R_1 = R_2$ , for  $R_3 = 6$  as well as  $\sigma_{\max}(\mathbf{P}_M \hat{\mathbf{W}})$ ; since the latter almost does not change, the lower bound is tight. In Figure 17, we see that there is a highest relative increase of the condition number around  $R_1 = R_2 = 32$ , which coincides with the point of the performance drop in Figure 15. Similar behaviour can be observed for the Salinas dataset on Figure 18.

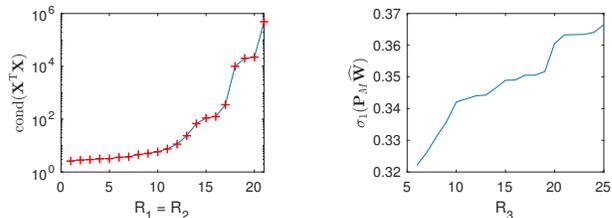


Fig. 18. Left:  $\log(\text{cond}\{\mathbf{X}^T \mathbf{X}\})$  depending on  $R_1 = R_2$  for  $R_3 = K_M$ ; right:  $\sigma_1(\mathbf{P}_M \hat{\mathbf{W}})$ , Salinas-A scene

All in all, based on the above examples, we can conclude that if we are in subregion (b), the  $R_3$  should be taken as large as possible ( $R_3 = K_M$ ), while in the subregion (a)  $R_1, R_2$  should be taken as large as possible while maintaining the condition number to a reasonable value.

#### F. Recovery of spectral signatures

Since correct recovery of spectral signatures is quite important for further processing of hyperspectral images, we would like to see whether SCOTT is able to do that. We consider the Indian Pines dataset, where groundtruth data (see Fig. 19) is available, splitting the image into 16 regions. We will consider three representative ranks:  $[40, 40, 6]$ ,  $[30, 30, 16]$ , and  $[24, 24, 25]$ , and compare them to STEREO ( $F = 100$ ).

We do not perform a proper hyperspectral unmixing, and compute the spectral signatures by averaging across the regions. We selected four representative signatures corresponding to materials 4, 7, 9 and 14, which are plotted in Figure VI-F. Note that materials 7 and 9 are scarce in the groundtruth SRI (resp. 28 and 20 pixels), whereas materials 4 and 14 are more abundant (resp. 237 and 1265 pixels).

In Figure 21 we plot relative errors of the reconstruction of spectra by different methods. As expected, for materials 7 and

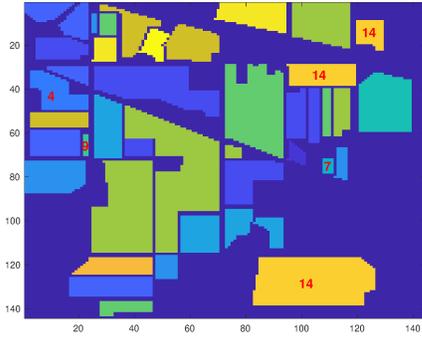


Fig. 19. Groundtruth image for Indian Pines dataset. Materials 4,7,9,14 are marked in red.

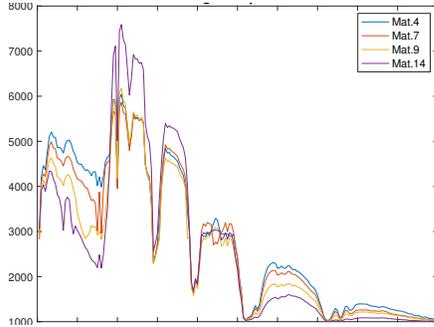


Fig. 20. Original spectral signature for materials 4,7,9 and 14

9, the discrepancy between the original spectra and the spectra obtained from estimated SRI is bigger than for materials 4 and 14. This can be explained by the scarcity of sources 7 and 9 compared to sources 4 and 14.

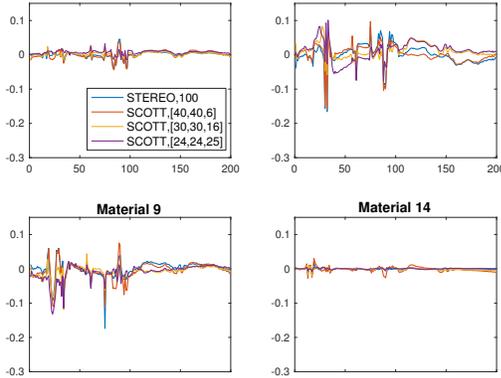


Fig. 21. Residual errors for the three considered ranks and four materials

In Figure 22, we have a closer look at the spectra at spectral bins 80 to 100. We can see that for abundant materials (4 and 14), all the algorithms estimate well the spectra. For the scarce materials it is important to choose the rank large enough, in particular  $R_3 = 16$  and  $R_3 = 25$  yield better reconstruction than  $R_3 = 6$ , and also than STEREO, even with  $F = 100$ .

## VII. CONCLUSION

In this paper, we proposed a novel coupled Tucker model for hyperspectral superresolution. We showed that the model is *recoverable*, that is, it allows for a unique recovery of

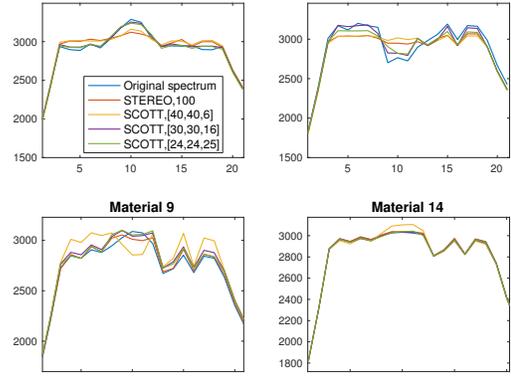


Fig. 22. Materials at spectral bins 80 to 100. Groundtruth (*black*), SCOTT (40, 40, 6) (*red*), SCOTT [30, 30, 16] (*yellow*), SCOTT (24, 24, 25) (*purple*), STEREO  $F = 100$  (*green*).

the SRI for a wide range of multilinear ranks. We proposed two very simple SVD-based algorithms can be used for the super-resolution problem, for known and unknown degradation operators, and for the case of pansharpening. The algorithms are very fast, but produce the results that are comparable with the CP-based approaches. Also, our algorithms are SVD-based, and thus do not have the drawbacks of the CPD (i.e., dependence on random initializations, ill-posedness of the CP approximation problem in  $\mathbb{R}$ ). This work opens new perspectives on using various tensor factorizations for hyperspectral super-resolution. Still several interesting questions remain, for example, how to enlarge the recoverability range for the multilinear rank. Estimating the multilinear rank of the Tucker decomposition still remains an open problem; the question of optimal splitting the data into non-overlapping subtensors in BSCOTT also needs to be further investigated.

## APPENDIX A SOLVING NORMAL EQUATIONS AS GENERALIZED SYLVESTER EQUATIONS

Equation (6) can be seen as a generalized Sylvester equation of the form

$$\mathbf{A}\hat{\mathbf{G}}\mathbf{B} + \mathbf{C}\hat{\mathbf{G}}\mathbf{D} = \mathbf{E}, \quad (19)$$

where  $\mathbf{G}$  is an unfolding of  $\hat{\mathbf{G}}$ .

We propose two options for converting (6) into (19). In the first case,  $\hat{\mathbf{G}} = \mathcal{G}^{(3)} \in \mathbb{R}^{R_1 R_2 \times R_3}$ ,

$$\mathbf{A} = \left( \mathbf{U}^\top \mathbf{P}_1^\top \mathbf{P}_1 \mathbf{U} \right) \boxtimes \left( \mathbf{V}^\top \mathbf{P}_2^\top \mathbf{P}_2 \mathbf{V} \right), \quad \mathbf{B} = \mathbf{I}_{R_3},$$

$$\mathbf{C} = \mathbf{I}_{R_1 R_2}, \quad \mathbf{D} = \lambda \left( \mathbf{W}^\top \mathbf{P}_M^\top \mathbf{P}_M \mathbf{W} \right),$$

and  $\mathbf{E} \in \mathbb{R}^{R_1 R_2 \times R_3}$  is a matricization of  $\mathbf{X}^\top \mathbf{z}$ .

In the second case,  $\hat{\mathbf{G}} = \mathcal{G}^{(1)\top} \in \mathbb{R}^{R_1 \times R_2 R_3}$ ,

$$\mathbf{A} = \mathbf{U}^\top \mathbf{P}_1^\top \mathbf{P}_1 \mathbf{U}, \quad \mathbf{B} = \mathbf{I}_{R_3} \boxtimes \left( \mathbf{V}^\top \mathbf{P}_2^\top \mathbf{P}_2 \mathbf{V} \right),$$

$$\mathbf{C} = \mathbf{I}_{R_1}, \quad \mathbf{D} = \lambda \left( \mathbf{W}^\top \mathbf{P}_M^\top \mathbf{P}_M \mathbf{W} \right) \boxtimes \mathbf{I}_{R_2},$$

and  $\mathbf{E} \in \mathbb{R}^{R_1 \times R_2 R_3}$  is a matricization of  $\mathbf{X}^\top \mathbf{z}$ .

The two options are equivalent and the fastest one is chosen according to the multilinear rank. As a rule of thumb, we decide to choose the first option in subregion (a) of Figure 1

and the second option in subregion (b). The complexity for solving the generalized Sylvester equation (19) is thus  $O(m^3 + n^3)$  flops for  $\widehat{G} \in \mathbb{R}^{m \times n}$  if fast solvers, such as Hessenberg-Schur or Bartels-Stewart methods [16], [17], are used.

## APPENDIX B DEGRADATION MATRICES

Here, we explain in details how the degradation matrices are constructed. For this appendix, we consider that  $P_1 = P_2$ . As in [8],  $P_1$  is constructed as  $P_1 = S_1 T_1$ , where  $T_1$  is a blurring matrix and  $S_1$  is a downsampling matrix.

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

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

Thus,  $T_1 \in \mathbb{R}^{I \times I}$  can be seen as

$$T_1 = \begin{bmatrix} \phi(\lfloor \frac{q}{2} \rfloor) & \dots & \phi(q) & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \phi(1) & & \ddots & & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \phi(q) & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \phi(1) & \dots & \phi(\lfloor \frac{q}{2} \rfloor) \end{bmatrix}.$$

The downsampling matrix  $S_1 \in \mathbb{R}^{I_H \times I}$ , with downsampling ratio  $d$ , is made of  $I_H$  independant rows such that for  $i = 1, \dots, I_H$ ,  $(S_1)_{i, 2+(i-1)d} = 1$  and the other coefficients are zeros.

The spectral degradation matrix  $P_M \in \mathbb{R}^{K_M \times K}$  is a selection-averaging matrix. Each row represents a spectral range in the MSI; coefficients are set to ones for common bands with the SRI, and zeros elsewhere. The coefficients are averaged per-row. Below, we give an example of a  $2 \times 6$  matrix:

$$\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

## REFERENCES

- [1] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 29–56, 2017.
- [2] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, 2015.
- [3] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, and C. Chen, "25 years of pansharpening: a critical review and new developments," *Signal and Image Process. for Remote Sens.*, pp. 533–548, 2011.
- [4] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012.
- [5] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.
- [6] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 33733388, 2015.

- [7] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, 2016.
- [8] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.
- [9] J. Hatvani, A. Basarab, J.-Y. Tourneret, M. Gyöngy, and D. Kouamé, "A Tensor Factorization Method for 3D Super-Resolution with Application to Dental CT," *IEEE Trans. Med. Imag.*, 2019, arxiv:1807.10027, doi:10.1109/TMI.2018.2883517.
- [10] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W. Ma, "Hyperspectral super-resolution: Combining low rank tensor and matrix structure," in *2018 25th IEEE International Conf. on Image Process. (ICIP)*, Oct 2018, pp. 3318–3322.
- [11] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [12] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Coupled tensor low-rank multilinear approximation for hyperspectral super-resolution," in *ICASSP*, 2019, available from <https://hal.archives-ouvertes.fr/hal-02025385>.
- [13] P. Comon, "Tensors: A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, 2014.
- [14] T.G. Kolda and B.W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [15] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. and Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [16] R.H. Bartels and G.W. Stewart, "Solution of the matrix equation  $AX+XB=C$ ," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [17] G. Golub, S. Nash, and C. Van Loan, "A Hessenberg-Schur method for the problem  $AX+XB=C$ ," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 909–913, 1979.
- [18] Y. Qi, P. Comon, and Lim L.-H., "Semialgebraic geometry of nonnegative tensor rank," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 4, pp. 1556–1580, 2016.
- [19] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," *IEEE Conf. on Comput. Vision and Pattern Recogn. (CVPR)*, p. 5344 5353, 2017.
- [20] A.J. Laub, *Matrix Analysis For Scientists And Engineers*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.
- [21] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 1, pp. 111–129, 2012.
- [22] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.
- [23] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, Available online.
- [24] "Grupo de inteligencia computacional (gic)," [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes), accessed: 2018-10-23.
- [25] "HSR via tensor decomposition (github repository)," [https://github.com/marhar19/HSR\\_via\\_tensor\\_decomposition](https://github.com/marhar19/HSR_via_tensor_decomposition), accessed: 2019-04-11.
- [26] P. Comon, "Tensors versus matrices, usefulness and unexpected properties," *IEEE Workshop on Statistical Signal Processing*, pp. 780–788, Sep 2009.
- [27] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SIAM J. Matrix Anal. Appl., Soc. Ind. Appl. Math.*, vol. 30, no. 3, pp. 1254–1279, 2008.
- [28] J. Brachat, P. Comon, B. Mourrain, and E.P. Tsigaridas, "Symmetric tensor decomposition," *17th European Signal Process. Conf. 2009*, pp. 525–529, Aug 2009.
- [29] P. Comon, "Tensor decompositions, state of the art and applications," in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002, available at arxiv:0905.0454.