Hyperspectral super-resolution with coupled Tucker approximation: Identifiability and SVD-based algorithms

Clémence Prévost, Konstantin Usevich, Pierre Comon, David Brie

HYPERSPECTRAL SUPER-RESOLUTION WITH COUPLED TUCKER APPROXIMATION: IDENTIFIABILITY AND SVD-BASED ALGORITHMS

C. Prévost[1], *K. Usevich*[1], *P. Comon*[2], *D. Brie*[1]

[1] CRAN, Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy, France
[2] GIPSA-Lab, Univ. Grenoble Alpes, CNRS, Saint-Martin d'Hères, France
clemence.prevost@univ-lorraine.fr, konstantin.usevich@univ-lorraine.fr
pierre.comon@gipsa-lab.grenoble-inp.fr, david.brie@univ-lorraine.fr

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

1. INTRODUCTION

The problem of *hyperspectral super-resolution* (HSR) [1] has recenlty 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 of this method is to recover a *super-resolution image* (SRI), which possesses both good spatial and spectral resolutions. This problem is closely related to hyperspectral pansharpening [2, 3], where HSI is fused with a panchromatic image (i.e. an "MSI" with only 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 linear mixing model widely used in hyperspectral image unmixing, 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 CP decomposition, the HSR is reformulated as a coupled CP (canonical polyadic) approximation. An alternating least squares algorithms (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 CP decomposition is identifiable, hence the HSI tensor can be uniquely recovered. This approach was also successfully used recently for a super-resolution problem in medical imaging [9]. Still, it has several drawbacks: for instance, the appropriate rank of the CP decomposition is not known *a priori* and may be unrelated to the number of endmembers; the rank can be also very large (esp. if images are treated as 2D arrays), which may affect computational complexity and convergence of the ALS iterations.

In many 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 alternating least squares 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. A key idea proposed in [10] is to use local approximations via blocking. 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 [8] and the second (blind) is inspired by [10]. Second, we show that, although the Tucker decomposition is not identifiable[1], surprisingly, the correct recovery of the SRI holds for a wide range of multilinear ranks. Our experiments on a number of semi-real and real examples, show that the proposed algorithms have a performance comparable to the ones of [8] and [10], but the computational cost is lower. Also, the algorithms can accurately reconstruct the spectral signatures, which is of prime importance for a further pro-

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

cessing of HSR image. Finally, we show that the proposed approach is applicable in the case of hyperspectral pansharpening (unlike [8], which requires the MSI to have at least two spectral bands).

This paper is organized as follows. In Section 2, we recall the HSR problem and the STEREO algorithm proposed in [8]. Section 3 contains our proposed coupled Tucker model and SVD-based algorithms (SCOTTand B-SCOTT) for tensor approximation. In Section 4 we prove our main identifiability result (recoverability) for the coupled Tucker model. Section 4 contains the numerical experiments.

Notation. In this paper we mainly follow [11] in what concerns the tensor notation (see also [12]). The following fonts are used: lowercase (a) or uppercase (I) plain font for scalars, boldface lowercase ($\boldsymbol{a}$) for vectors, uppercase boldface ($\boldsymbol{A}$) for matrices, and calligraphic ($\boldsymbol{\mathcal{A}}$) 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 $\mathcal{A}_{i_1,\ldots,i_N}$ respectively. $\mathbb{R}$ stands for the real line, $\mathbb{C}$ for the set of complex numbers.

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

We use $\mathrm{vec}\{\cdot\}$ for the standard column-major vectorization of a tensor or a matrix. Operator $\bullet_p$ denotes contraction on the pth index of a tensor; when contracted with a matrix, it is understood that summation is always performed on the second index of the matrix. For instance, $[\boldsymbol{\mathcal{A}} \bullet_1 \boldsymbol{M}]_{ijk} = \sum_\ell \mathcal{A}_{\ell jk} M_{i\ell}$. For a tensor $\boldsymbol{\mathcal{G}}$ and matrices $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$, the following shorthand notation is used

$$[\![\boldsymbol{\mathcal{G}}; \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}]\!] = \boldsymbol{\mathcal{G}} \underset{1}{\bullet} \boldsymbol{U} \underset{2}{\bullet} \boldsymbol{V} \underset{3}{\bullet} \boldsymbol{W}.$$

For matrices $\boldsymbol{A} \in \mathbb{R}^{I\times F}$, $\boldsymbol{B} \in \mathbb{R}^{J\times F}$, $\boldsymbol{C} \in \mathbb{R}^{K\times F}$, we will use a shorthand notation for the polyadic decomposition

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

where $\boldsymbol{\mathcal{I}}_F \in \mathbb{R}^{F\times F\times F}$ is a diagonal tensor of ones. For a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$, its first unfolding is denoted by $\boldsymbol{Y}^{(1)} \in \mathbb{R}^{JK\times I}$. By $\mathrm{tSVD}_R(\boldsymbol{X})$ we denote a matrix containing R leading right singular vectors of the matrix $\boldsymbol{X}$.

2. HYPERSPECTRAL DATA FUSION PROBLEM

2.1. Problem statement and degradation model

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

In this paper, we adopt the following degradation model, that can be compactly written as contraction of SRI with degradation matrices:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_M &= \boldsymbol{\mathcal{Y}} \bullet_3 \boldsymbol{P_M} + \boldsymbol{\mathcal{E}}_M, \\ \boldsymbol{\mathcal{Y}}_H &= \boldsymbol{\mathcal{Y}} \bullet_1 \boldsymbol{P_1} \bullet_2 \boldsymbol{P_2} + \boldsymbol{\mathcal{E}}_H, \end{cases} \tag{1}$$

where $\boldsymbol{\mathcal{E}}_M$ and $\boldsymbol{\mathcal{E}}_H$ denote the noise terms, $\boldsymbol{P_M} \in \mathbb{R}^{K_M\times K}$ is the spectral degradation matrices (for example, a selection-averaging matrix), $K_M < K$, and $\boldsymbol{P_1} \in \mathbb{R}^{I_H\times I}$, $\boldsymbol{P_2} \in \mathbb{R}^{J_H\times J}$, $I_H < I$, $J_H < J$, are the spatial degradation matrices[2], i.e. we assume (for simplicity) that the spatial degradation is separable; this is a valid assumption, for example, for the commonly accepted Wald's protocol [13], which uses Gaussian blurring and downsampling in both spatial dimensions. In this paper we consider only the case when the degradation matrices $\boldsymbol{P_1}, \boldsymbol{P_2}, \boldsymbol{P_M}$ are known.

2.2. CP-based approach (STEREO)

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

$$\underset{\widehat{\boldsymbol{A}},\widehat{\boldsymbol{B}},\widehat{\boldsymbol{C}}}{\text{minimize}} \quad f_{CP}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}), \tag{2}$$

where $f_{CP}(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}) =$

$$\|\boldsymbol{\mathcal{Y}}_H - [\![\boldsymbol{P_1}\widehat{\boldsymbol{A}}, \boldsymbol{P_2}\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}]\!]\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_M - [\![\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \boldsymbol{P_M}\widehat{\boldsymbol{C}}]\!]\|_F^2,$$

which is a coupled CP approximation problem. For the case when there is no noise $\mathcal{E}_H, \mathcal{E}_M = 0$, 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 more details.

To solve (2), an alternating optimization algorithm is proposed in [8], called STEREO (Super-Resolution Tensor Reconstruction).

The updates of the factor matrices in Algorithm 1 can be computed by using efficient solvers for the (generalized) Sylvester equation [14], [15]. For example, the total cost (in flops) of one iteration in Algorithm 1 becomes (updating A, B, C)

[2]Typically, it holds that $I_H J_H \ll IJ$ and $K_M \ll K$, meaning that MSI has a finer spectral resolution w.r.t HSI, and HSI has a finer spatial resolution than MSI.

input : $\boldsymbol{A}_0 \in \mathbb{R}^{I\times F}$, $\boldsymbol{B}_0 \in \mathbb{R}^{J\times F}$, $\boldsymbol{C}_0 \in \mathbb{R}^{K\times F}$, F
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
for $k = 1 : n$ **do**
$\quad \boldsymbol{A}_k \leftarrow \underset{\boldsymbol{A}}{\operatorname{argmin}}\, f_{CP}(\boldsymbol{A}, \boldsymbol{B}_{k-1}, \boldsymbol{C}_{k-1})$,
$\quad \boldsymbol{B}_k \leftarrow \underset{\boldsymbol{B}}{\operatorname{argmin}}\, f_{CP}(\boldsymbol{A}_k, \boldsymbol{B}, \boldsymbol{C}_{k-1})$,
$\quad \boldsymbol{C}_k \leftarrow \underset{\boldsymbol{C}}{\operatorname{argmin}}\, f_{CP}(\boldsymbol{A}_k, \boldsymbol{B}_k, \boldsymbol{C})$,
end
$\widehat{\boldsymbol{A}} \leftarrow \boldsymbol{A}_n, \widehat{\boldsymbol{B}} \leftarrow \boldsymbol{B}_n, \widehat{\boldsymbol{C}} \leftarrow \boldsymbol{C}_n, \widehat{\boldsymbol{\mathcal{Y}}} \leftarrow [\![\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}]\!]$.
Algorithm 1: STEREO

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

For more details on these equations, see[3] [8, App. E].

The initial values in Algorithm 1 are chosen as

$$\begin{aligned}\left(\boldsymbol{A}_0, \boldsymbol{B}_0, \tilde{\boldsymbol{C}}_0\right) &= \text{CPD}_F\left(\boldsymbol{\mathcal{Y}}_M\right),\\ \boldsymbol{Y_M}^{(3)} &= \left(\boldsymbol{P_1 A_0} \odot \boldsymbol{P_2 B_0}\right) \boldsymbol{C}_0^{\mathsf{T}},\\ \tilde{\boldsymbol{C}}_0 &= \boldsymbol{P_M C_0},\end{aligned}$$

where CPD_F stands for a rank-F CP approximations and the last two equations are solved in the least squares sense.

2.3. 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 $\boldsymbol{P_1}$,$\boldsymbol{P_2}$ are unknown, proposed in [8] and [10]. The first solution, called blind STEREO was to consider the following coupled CP approximation problem:

$$\min \|\boldsymbol{\mathcal{Y}}_H - [\![\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widehat{\boldsymbol{C}}]\!]\|_F^2 + \lambda \|\boldsymbol{\mathcal{Y}}_M - [\![\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \boldsymbol{P_M}\widehat{\boldsymbol{C}}]\!]\|_F^2,$$

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

In [10], an alternative approach was proposed, that uses the CP approximation of $\boldsymbol{\mathcal{Y}}_M$ to retrieve the factors $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}})$, an SVD of the unfolding of $\boldsymbol{\mathcal{Y}}_H$, and a least squares problem to retrieve the third factor $\widehat{\boldsymbol{C}}$. This approach, which does not necessary need separability of the spatial degradation operation, is summarized in Algorithm 2.

As noted in [10], the correct recovery of a low-rank tensor $\boldsymbol{\mathcal{Y}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ is only if $\text{rank}\{\boldsymbol{C}\} = R$ does not exceed the number of spectral bands in MSI, because the rank of $\widehat{\boldsymbol{C}}$ is bounded by K_M. 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 will be active in a smaller black). 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.

[3]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 [14].

input : $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I\times J\times K_M}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H\times J_H\times K}$, R, F, $\boldsymbol{P}_M$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
Compute CP approximation:
$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widetilde{\boldsymbol{C}}) = \text{CPD}_F(\boldsymbol{\mathcal{Y}}_M)$
$\boldsymbol{Z} \leftarrow \text{tSVD}_R\left(\boldsymbol{Y_H}^{(3)}\right)$
$\widehat{\boldsymbol{C}} \leftarrow \boldsymbol{Z}(\boldsymbol{P}_M\boldsymbol{Z})^{\dagger}\widetilde{\boldsymbol{C}}$
$\widehat{\boldsymbol{\mathcal{Y}}} \leftarrow [\![\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}]\!]$.
Algorithm 2: Hybrid algorithm of [10]

input : $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I\times J\times K_M}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H\times J_H\times K}$, R, F, $\boldsymbol{P}_M$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
Split each $\boldsymbol{\mathcal{Y}}_M$ and $\boldsymbol{\mathcal{Y}}_H$ in L subblocks along the spatial dimension (the subblocks correspond to each other).
for $k = 1 : n$ **do**
$\quad$ Apply Algorithm 2 to each subblock, and store the result in the corresponding subblock of $\widehat{\boldsymbol{\mathcal{Y}}}$.
end
Algorithm 3: SCUBA

3. TUCKER-BASED DATA FUSION

3.1. Model and approximation problem

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

With these notations, Equation (1) becomes

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

The HSR formulation is thus

$$\underset{\widehat{\boldsymbol{\mathcal{G}}}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}}{\text{minimize}}\ f_T(\widehat{\boldsymbol{\mathcal{G}}}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}), \quad \text{where} \tag{3}$$

$$\begin{aligned}f_T(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}, \widehat{\boldsymbol{\mathcal{G}}}) =& \|\boldsymbol{\mathcal{Y}}_H - [\![\widehat{\boldsymbol{\mathcal{G}}}; \boldsymbol{P_1}\widehat{\boldsymbol{U}}, \boldsymbol{P_2}\widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}]\!]\|_F^2\\ &+\lambda \|\boldsymbol{\mathcal{Y}}_M - [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \boldsymbol{P_M}\widehat{\boldsymbol{W}}]\!]\|_F^2.\end{aligned} \tag{4}$$

3.2. An SVD-based algorithm

A suboptimal SVD-based solution to problem (3) can be found by a simple method that is similat to the well-known high-order SVD (HOSVD) of tensors [16]. This algorithm, named as SCOTT(Super-resolution based on COupled Tucker Tensor approximation), is given in Algorithm 4.

input : $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I\times J\times K_M}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H\times J_H\times K}$, $\boldsymbol{R}$, $\boldsymbol{P}_M$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
1. $\widehat{\boldsymbol{U}} \leftarrow \text{tSVD}_{R_1}\left(\boldsymbol{Y_M}^{(1)}\right)$, $\widehat{\boldsymbol{V}} \leftarrow \text{tSVD}_{R_2}\left(\boldsymbol{Y_M}^{(2)}\right)$, $\widehat{\boldsymbol{W}} \leftarrow \text{tSVD}_{R_3}\left(\boldsymbol{Y_H}^{(3)}\right)$,
2. $\widehat{\boldsymbol{\mathcal{G}}} \leftarrow \underset{\boldsymbol{\mathcal{G}}}{\text{argmin}}\ f_T(\boldsymbol{\mathcal{G}}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}})$
3. $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}]\!]$.

Algorithm 4: SCOTT

The least squares subproblem in Algorithm 4

$$\underbrace{\begin{bmatrix} \widehat{\boldsymbol{W}} \boxtimes \boldsymbol{P_2}\widehat{\boldsymbol{V}} \boxtimes \boldsymbol{P_1}\widehat{\boldsymbol{U}} \\ \sqrt{\lambda}\boldsymbol{P_M}\widehat{\boldsymbol{W}} \boxtimes \widehat{\boldsymbol{V}} \boxtimes \widehat{\boldsymbol{U}} \end{bmatrix}}_{\boldsymbol{X}} \text{vec}\{\widehat{\boldsymbol{\mathcal{G}}}\} \approx \underbrace{\begin{bmatrix} \text{vec}\{\boldsymbol{\mathcal{Y}}_H\} \\ \sqrt{\lambda}\,\text{vec}\{\boldsymbol{\mathcal{Y}}_M\} \end{bmatrix}}_{\boldsymbol{z}}$$

can be solved through normal equations of the form

$$\left(\boldsymbol{X}^\mathsf{T}\boldsymbol{X}\right)\text{vec}\{\widehat{\boldsymbol{\mathcal{G}}}\} = \boldsymbol{X}^\mathsf{T}\boldsymbol{z},$$

where the matrix on the left-hand side is

$$\boldsymbol{X}^\mathsf{T}\boldsymbol{X} = \boldsymbol{I}_{R_3} \boxtimes (\widehat{\boldsymbol{V}}^\mathsf{T}\boldsymbol{P}_2^\mathsf{T}\boldsymbol{P}_2\widehat{\boldsymbol{V}}) \boxtimes (\widehat{\boldsymbol{U}}^\mathsf{T}\boldsymbol{P}_1^\mathsf{T}\boldsymbol{P}_1\widehat{\boldsymbol{U}}) + \lambda(\widehat{\boldsymbol{W}}^\mathsf{T}\boldsymbol{P}_M^\mathsf{T}\boldsymbol{P}_M\widehat{\boldsymbol{W}}) \boxtimes \boldsymbol{I}_{R_1R_2}$$

The normal equations can be viewed as a Sylvester equation and (as in the case of STEREO) efficient solvers can be used. Thus the total cost of one iteration in SCOTT(in flops) becomes

- $O(\min(R_1, R_2)IJK_M + R_3I_HJ_HK)$ for computing the truncated SVDs and computing $\boldsymbol{X}^\mathsf{T}\boldsymbol{z}$;
- $O(R_3^3 + (R_1R_2)^3)$ for solving the Sylvester equation.

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

3.3. An algorithm for unknown spatial degradation

Here, we show that is also possible to develop a blind SVD-based algorithm, in the same spirit as Algorithm 2. The algorithm does not need knowledge of $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and is based on the HOSVD of the MSI tensor.

input : $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I\times J\times K_M}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H\times J_H\times K}$, (R_1, R_2, R_3), $\boldsymbol{P}_M$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
1. Compute the (R_1, R_2, R_3) HOSVD Tucker approximation of $\boldsymbol{\mathcal{Y}}_M$

$$[\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widetilde{\boldsymbol{W}}]\!] \overset{\text{HOSVD}}{\approx} \boldsymbol{\mathcal{Y}}_M$$

2. $\boldsymbol{Z} \leftarrow \text{tSVD}_{R_3}\left(\boldsymbol{Y_H}^{(3)}\right)$
3. $\widehat{\boldsymbol{W}} \leftarrow \boldsymbol{Z}(P_M\boldsymbol{Z})^\dagger\widetilde{\boldsymbol{W}}$
4. $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}]\!]$.

Algorithm 5: Blind version of SCOTT

The total computational complexity (in flops) of Algorithm 5 is

$$O(\min(R_1, R_2)IJK_M + R_3I_HJ_HK)$$

and is dominated by the costs of the truncated SVD, because step 3 is very cheap.

Finally, similarly to SCUBA, we can use a block version of Algorithm 5, which we call B-SCOTT(that is "Blind SCOTT"). There is no confusion, as Algorithm 5 is a special case of Algorithm 6 in the case of 1 block.

input : $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I\times J\times K_M}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H\times J_H\times K}$, (R_1, R_2, R_3), $\boldsymbol{P}_M$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
Partition each $\boldsymbol{\mathcal{Y}}_M$ and $\boldsymbol{\mathcal{Y}}_H$ in L subblocks along the spatial dimension (the subblocks correspond to each other).
for $k = 1 : n$ **do**
  Apply Algorithm 5 to each subblock, and store the result in the corresponding subblock of $\widehat{\boldsymbol{\mathcal{Y}}}$.
end

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

4. RECOVERABILITY OF THE TUCKER MODEL

In this subsection, we establish a generic uniqueness result for the tensor recovery in the coupled Tucker model.

Theorem 4.1. *Assume that $\boldsymbol{P_1} \in \mathbb{R}^{I_H\times I}$, $\boldsymbol{P_2} \in \mathbb{R}^{J_H\times J}$, and $\boldsymbol{P_M} \in \mathbb{R}^{K_M\times K}$ are fixed full row-rank matrices. Let*

$$\boldsymbol{\mathcal{Y}} = [\![\boldsymbol{\mathcal{G}}; \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}]\!], \tag{5}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1\times R_2\times R_3}$, $R_1 \le I$, $R_2 \le J$, $R_3 \le K$, and $\boldsymbol{U} \in \mathbb{R}^{I\times R_1}$, $\boldsymbol{V} \in \mathbb{R}^{J\times R_2}$, $\boldsymbol{W} \in \mathbb{R}^{K\times R_3}$ are random matrices, distributed according to an absolutely continuous probability distribution. We also assume that $\boldsymbol{\mathcal{E}}_M, \boldsymbol{\mathcal{E}}_H = \boldsymbol{0}$ in (1).

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

$$\begin{cases} R_1 \leq \min(R_3, K_M) R_2, \\ R_2 \leq \min(R_3, K_M) R_1 \\ R_3 \leq \min(R_1, I_H) \min(R_2, J_H), \end{cases} \tag{6}$$

then with probability 1 *there exists a unique tensor* $\widehat{\mathcal{Y}}$ *such that* $\widehat{\mathcal{Y}}_M = \mathcal{Y}_M$ *and* $\widehat{\mathcal{Y}}_H = \mathcal{Y}_H$.

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 an continuum of* $\widehat{\mathcal{Y}}$ *such that* $\widehat{\mathcal{Y}}_M = \mathcal{Y}_M$ *and* $\widehat{\mathcal{Y}}_H = \mathcal{Y}_H$*; in fact,* $\|\widehat{\mathcal{Y}} - \mathcal{Y}\|$ *can be arbitrary large.*

Corollary 4.2. *If* $\mathcal{Y}$ *is as in* (5)*, such that the conditions of Theorem 4.1 (part 1.) hold., then any minimizer of* (3) *recovers* $\mathcal{Y}$*, i.e.*

$$\mathcal{Y} = [\![\widehat{\mathcal{G}}; \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}]\!].$$

In addition, it follows that Algorithm 4 (SCOTT) recovers $\mathcal{Y}$ *for all cases of recoverability in Theorem 4.1, and Algorithm 5 (B-SCOTT) if* $R_3 \leq K_M$.

Before proceeding to the proof of Theorem 4.1, we illustrate the statement of Theorem 4.1 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: identifiable and non-identifiable. The hatched area corresponds to the parameters where condition (6) is not satisfied .

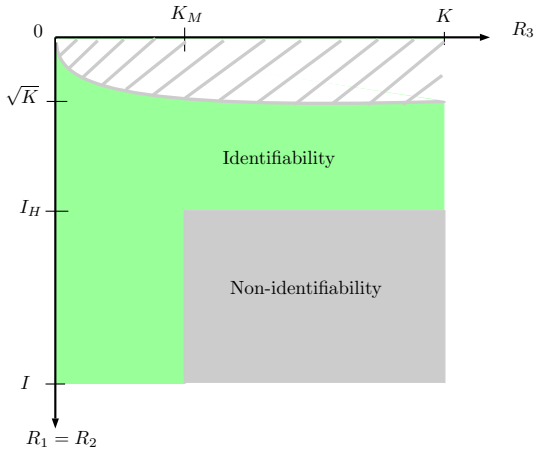


Fig. 1. Identifiability region depending on R_1 and R_3

Proof of Theorem 4.1. First, without loss of generality, we can replace $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_M$ with

$$\widetilde{\boldsymbol{P}}_1 = \begin{bmatrix} \boldsymbol{I}_{R_1} & \boldsymbol{0} \end{bmatrix}, \widetilde{\boldsymbol{P}}_2 = \begin{bmatrix} \boldsymbol{I}_{R_2} & \boldsymbol{0} \end{bmatrix}, \widetilde{\boldsymbol{P}}_M = \begin{bmatrix} \boldsymbol{I}_{R_3} & \boldsymbol{0} \end{bmatrix}. \tag{7}$$

Indeed, let us explain why it is so, for example for $\boldsymbol{P}_1 \in \mathbb{R}^{R_1 \times I}$. There exists a nonsingular matrix[4] $\boldsymbol{T}$ such that

$$\boldsymbol{P}_1 \boldsymbol{T} = \begin{bmatrix} \boldsymbol{I}_{R_1} & \boldsymbol{0} \end{bmatrix}.$$

[4]For example, $\boldsymbol{T} = [\boldsymbol{P}_1^\dagger \; \boldsymbol{F}]$, where $\boldsymbol{F} \in \mathbb{R}^{I \times (I - R_1)}$, $\boldsymbol{P}_1 \boldsymbol{F} = \boldsymbol{0}$.

If we take $\widetilde{\boldsymbol{U}} = \boldsymbol{T}^{-1} \boldsymbol{U}$ then we have

$$\boldsymbol{P}_1 \boldsymbol{U} = \widetilde{\boldsymbol{P}}_1 \widetilde{\boldsymbol{U}}.$$

Note that a nonsingular transformation preserves absolute continuity of the distribution; hence $\boldsymbol{U}$ has an absolutely continuous distribution if and only if $\widetilde{\boldsymbol{U}}$ has one.

The rest of the proof uses some basic real algebraic geometry and basic properties of semialgebraic sets. For a necessary background, we refer the reader to the introduction and references in [17]. In what follows, for simplicity, we will assume $\widetilde{\boldsymbol{P}}_1 = \boldsymbol{P}_1$, $\widetilde{\boldsymbol{P}}_2 = \boldsymbol{P}_2$, $\widetilde{\boldsymbol{P}}_3 = \boldsymbol{P}_3$. Next, we show that nonunique cases appear only on a set of measure zero in the parameter space $\{(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \mathcal{G})\} \in \mathscr{F}_1$

$$\mathscr{F}_1 = \mathbb{R}^{I \times R_1} \times \mathbb{R}^{J \times R_2} \times \mathbb{R}^{K \times R_3} \times \mathbb{R}^{R_1 \times R_2 \times R_3}.$$

For convenience, we will switch to a non-ambiguous parameterization. Consider the parameter space

$$\mathscr{F}_2 = \mathbb{R}^{(I\text{-}R_1) \times R_1} \times \mathbb{R}^{(J\text{-}R_2) \times R_2} \times \mathbb{R}^{(K\text{-}R_3) \times R_3} \times \mathbb{R}^{R_1 \times R_2 \times R_3}.$$

Take $\widetilde{\mathscr{F}_1} \subset \mathscr{F}_1$ such that $(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \mathcal{G}) \in \widetilde{\mathscr{F}_1}$ iff

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{bmatrix}, \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \end{bmatrix}, \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 \\ \boldsymbol{W}_2 \end{bmatrix},$$

and $\boldsymbol{U}_1 \in \mathbb{R}^{R_1 \times R_1}$, $\boldsymbol{V}_1 \in \mathbb{R}^{R_2 \times R_2}$, $\boldsymbol{W}_1 \in \mathbb{R}^{R_3 \times R_3}$ are nonsingular; thus we exclude a set $\mathscr{F}_1 \setminus \widetilde{\mathscr{F}_1}$ of measure zero. Consider the surjective reparameterization map

$$\begin{aligned} &\mathscr{R} : \widetilde{\mathscr{F}_1} \to \mathscr{F}_2 \\ &(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \mathcal{G}) \mapsto \left(\boldsymbol{U}_2 \boldsymbol{U}_1^{-1}, \boldsymbol{V}_2 \boldsymbol{V}_1^{-1}, \boldsymbol{W}_2 \boldsymbol{W}_1^{-1}, [\![\mathcal{G}; \boldsymbol{U}_1, \boldsymbol{V}_1, \boldsymbol{W}_1]\!]\right) \end{aligned}$$

The set $\mathscr{F}_2$ can be injectively mapped to $\mathbb{R}^{I \times J \times K}$

$$\begin{aligned} \mathscr{T} : &\mathscr{F}_2 \to \mathbb{R}^{I \times J \times K} \\ &(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \mathcal{S}) \mapsto \mathcal{Y} = [\![\mathcal{S}; \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{A} \end{bmatrix}, \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{B} \end{bmatrix}, \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{C} \end{bmatrix}]\!]. \end{aligned} \tag{8}$$

In fact, the mapping $\mathscr{T}$ is nothing more than a construction of the tensor from parameters, i.e.

$$\mathscr{T}(\mathscr{R}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \mathcal{G})) = [\![\mathcal{S}; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!] = [\![\mathcal{G}; \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}]\!];$$

Hence the core tensor $\mathcal{S}$ is just the core tensor $\mathcal{G}$ transformed in a new basis.

Now assume that $\mathcal{Y} = \mathscr{T}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \mathcal{S})$, and we would like to see when we can reconstruct $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \mathcal{S})$ uniquely from $\mathcal{Y}_H$ and $\mathcal{Y}_M$.

Proof of 2. Consider the case $R_1 > I_H$ and $R_3 > K_M$ (the case $R_2 > J_H$ and $R_3 > K_M$ is analogous). Then the tensors $(\mathcal{Y}_M)$ and $(\mathcal{Y}_H)$ do not depend on the values $\mathcal{S}_{i,:,:,k}$ for $i > I_H$ and $j > I$. Hence there is an infinite set of $\widehat{\mathcal{S}}$ such that $\mathcal{Y}_M = (\mathscr{T}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \widehat{\mathcal{S}}))_M$ and $\mathcal{Y}_H = (\mathscr{T}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \widehat{\mathcal{S}}))_H$. Since $\mathscr{R}^{-1} \mathscr{F}_2 = \widetilde{\mathscr{F}_1}$, the nonuniqueness occurs for all $(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \mathcal{G}) \in \widetilde{\mathscr{F}_1}$, i.e. with probability 1.

Proof of 1. Now let us prove the identifiability cases $R_3 \leq K_M$ or $(R_1, R_2) \leq (I_H, J_H)$. First, note that

- If $R_3 \leq K_M$, we have $\boldsymbol{\mathcal{S}} = (\boldsymbol{\mathcal{Y}}_M)_{1:R_1,1:R_2,1:R_3}$.
- If $(R_1, R_2) \leq (I_H, J_H)$, we have $\boldsymbol{\mathcal{S}} = (\boldsymbol{\mathcal{Y}}_H)_{1:R_1,1:R_2,1:R_3}$,

hence $\boldsymbol{\mathcal{S}}$ can be uniquely recovered from $\boldsymbol{\mathcal{Y}}_M$ and $\boldsymbol{\mathcal{Y}}_H$.

Next, we consider the question of recovery of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ from unfoldings. Consider reconstruction of $\boldsymbol{A}$ from the first unfolding of $\boldsymbol{\mathcal{Y}}_M$. Take the subtensor

$$\boldsymbol{\mathcal{Z}} = (\boldsymbol{\mathcal{Y}}_M)_{:,1:R_2,1:\min(R_3,K_M)}$$

and the subtensor $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{S}})_{:,:,1:\min(R_3,K_M)}$ of the core tensor $\boldsymbol{\mathcal{S}}$. Then the first unfolding $\boldsymbol{Z}^{(1)} \in \mathbb{R}^{R_2 \min(R_3,K_M) \times I}$ (which is a submatrix of $\boldsymbol{Y_M}^{(1)}$) has a low-rank factorization

$$\boldsymbol{Z}^{(1)} = \boldsymbol{X}^{(1)} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}^{\mathsf{T}} \end{bmatrix}.$$

hence $\boldsymbol{A}$ can be uniquely determined from $\boldsymbol{Z}^{(1)}$ if $\boldsymbol{X}^{(1)}$ has full-column rank. Note that if $R_1 \leq \min(R_3, K_M)$, then the set of $\boldsymbol{\mathcal{S}}$ such that $\boldsymbol{X}^{(1)}$ is not of full column rank is a semialgebraic subset of $\mathbb{R}^{R_1 \times R_2 \times R_3}$ of measure zero[5].

Proceeding similarly for $\boldsymbol{B}$ and $\boldsymbol{C}$, we get that if (6) holds, then there exists a semialgebraic set $\mathscr{F}_3 \subset \mathscr{F}_2$ of measure zero such that for $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\mathcal{S}}) \in \mathscr{F}_2 \setminus \mathscr{F}_3$ there is a perfect recovery of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\mathcal{S}})$ and $\boldsymbol{\mathcal{Y}}$.

By the semialgebraic version of Sard's theorem [17, Lemma 2.1], the preimage $\mathscr{R}^{-1}(\mathscr{F}_3)$ is also of measure zero, hence the uniqueness of recovery holds for $(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \boldsymbol{\mathcal{G}}) \in \widetilde{\mathscr{F}_1} \setminus \mathscr{R}^{-1}(\mathscr{F}_3)$, i.e., with probability 1. □

Remark 4.3. *In the proof of Theorem 4.1 it was shown that for the recoverability of the coupled model we can assume that the degradation operators are given in* (7)*. In that case, the degraded tensors $\boldsymbol{\mathcal{Y}}_M$ and $\boldsymbol{\mathcal{Y}}_H$ are just the subtensors (slabs)* i.e.

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

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

5. EXPERIMENTS

All simulations were run on a Macintosh computer with 2.4 GHz Intel Core i5 and 8GB RAM. The code is implemented in MATLAB. For basic tensor operations we used TensorLab 3.0 [18]. The results are reproducible and the codes are available online at

`https://github.com/cprevost4/HSR_Tucker`.

[5] i.e., a semialgebraic subset whose dimension is smaller than the dimension of the ambient space.

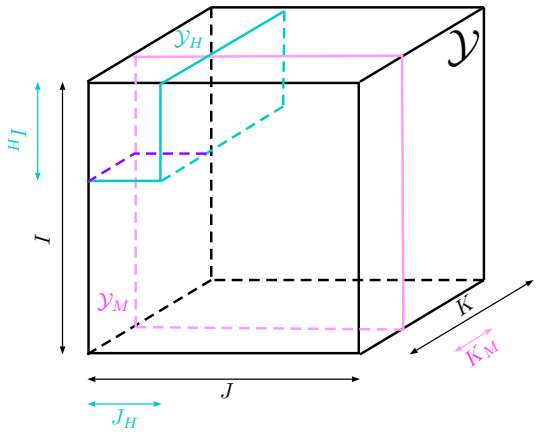


Fig. 2. Recovery of $\boldsymbol{\mathcal{Y}}$ (in black) from $\boldsymbol{\mathcal{Y}}_{:,:,1:K_M}$ (pink), $\boldsymbol{\mathcal{Y}}_{1:I_H,1:J_H,:}$ (blue)

As for the experimental setup, we follow [8]. The main performance metric used in comparisons is *reconstruction Signal-to-Noise ratio* (R-SNR) introduced in [3]

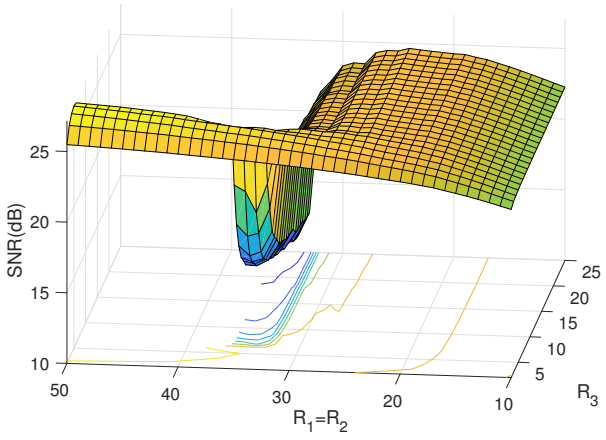
$$\text{R-SNR} = 10\log_{10}\left(\frac{\|\boldsymbol{\mathcal{Y}}\|_F^2}{\|\widehat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}\|_F^2}\right).$$

In all the examples, as in [8], the bands corresponding to water absorption are removed. For all the experiments (except the real wood data), the degradation matrices $\boldsymbol{P_1}$, $\boldsymbol{P_2}$ are generated following Wald's protocol, and the downsampling factor is chosen to be 4, see The matrix $\boldsymbol{P_M}$ is a selection-averaging matrix that splits the spectral range into parts according to the specification of the selected sensor.

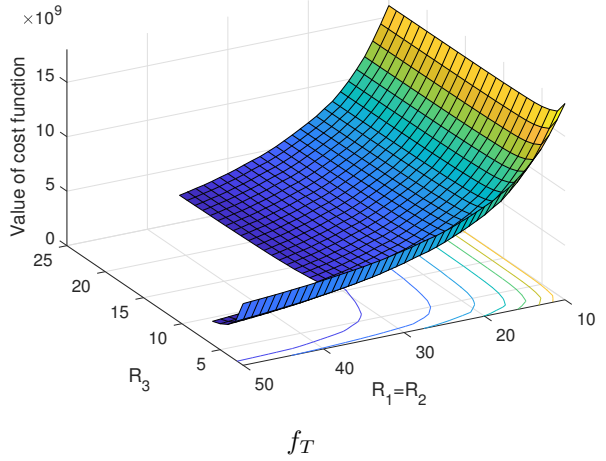
5.1. Choice of multilinear ranks

In this subsection, we have a closer look at the Indian Pines dataset, available at [19], that was acquired by a 224-band AVIRIS sensor. In this case, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{144\times144\times200}$ (after removing the bands corresponding to water absorption), $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{144\times144\times6}$ and $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{36\times36\times200}$. The spectral degradation is chosen according to the specifications of the LANDSAT sensor. We analyze the performance of SCOTT(R-SNR and the value of the cost function f_T defined in (4)) w.r.t. the multilinear rank employed in the Tucker decomposition.

Figure 3 shows the SNR and cost function value for $R_1 = R_2$ varying in the range of $[10 : 50]$ and R_3 varying in the range of $[2 : 25]$ for which the identifiability condition holds (see Section 4 and Fig. 1). While the cost function decreases as R_1 and R_3 increase, the reconstruction error (given by R-SNR) behaves differently for different identifiability subregions in Fig. 1. In the case $R_3 \leq K_H$, the R-SNR increases when R_1 grows, and R-SNR is close to maximal when $R_3 = K_H$. But, in the case $R_1 = R_2 \leq I_H$ and growing R_3, the situation is different: the R-SNR varies slowly with increasing R_3, but is more sensitive to the choice of $R_1 = R_2$.

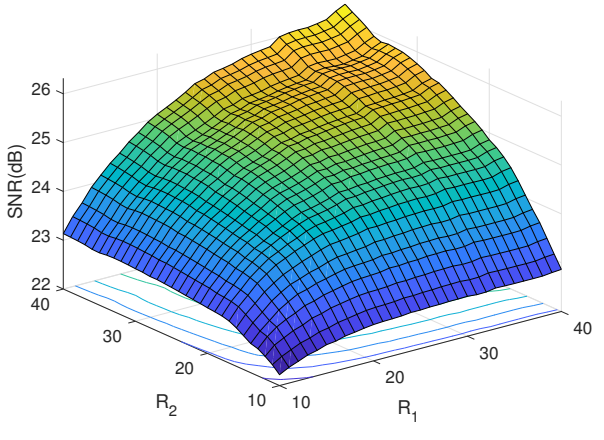


R-SNR

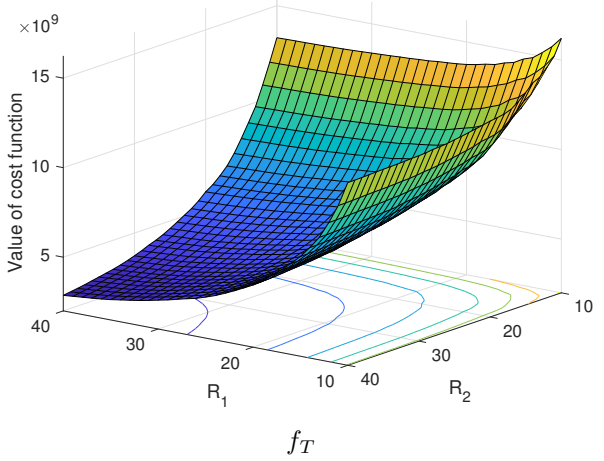


f_T

Fig. 3. R-SNR and f_T as a function of $R_1 = R_2$ and R_3



R-SNR



f_T

Fig. 4. R-SNR and f_T as a function of R_1 and R_2, $R_3 = 6$

In particular the overfitting (i.e. sharp decrase in R-SNR) is observed when $R_1 = R_2$ is close to I_H.

We also performed the optimisation of (3) with the structured data fusion framework implemented in Tensorlab [18]. In our experiment, for all the values in Fig. 3, the optimization procedure stopped in a few iterations and did not improve the cost function. This result seems reasonable, because it is known [16] that SCOTTgives a very good solution for a low-rank Tucker approximation. In what follows, we use only SVD-based algorithms for our coupled model.

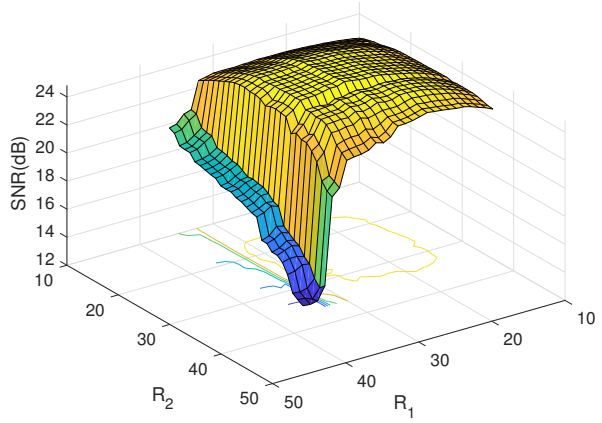
We conclude this subsection by testing the case of fixed R_3 and varying R_1 and R_2. In Figure 4, we fix $R_3 = 6$ and report the results for SCOTTfor R_1 and R_2 in a range of $[10 : 40]$. In Figure 5, we report the same results for $R_3 = 16$.

We observe different behaviour in the two different cases. In the case $R_3 = 6$ (see Fig. 4), the results are rather symmetric, whereas in Fig. 5 we see that the overfitting is mainly due to R_1 being close to I_H. Still, the best value is close to the diagonal $R_1 = R_2$, therefore in the following experiments we will only consider the case $R_1 = R_2$.

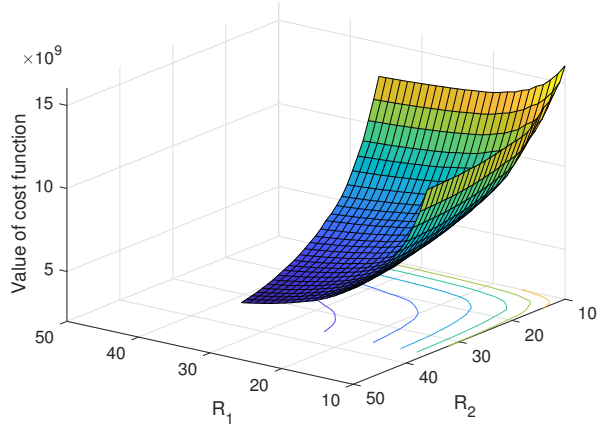
5.2. Comparison with other algorithms

In this subsection, we compare the performance of SCOTT(for a set of representative ranks), B-SCOTT(without partition into blocks), STEREO, and HySure [6]. In addition to R-SNR, we consider different metrics from [3] described below:

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



R-SNR



Cost value

Fig. 5. R-SNR and f_T as a function of R_1 and R_2, $R_3 = 16$

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

$$\text{SAM} = \frac{180}{\pi}\frac{1}{IJ}\sum_{n=1}^{IJ}\arccos\left(\frac{{\boldsymbol{Y}^{(3)}}^{\mathsf{T}}_{n,:}\,\widehat{\boldsymbol{Y}}^{(3)}_{n,:}}{\|{\boldsymbol{Y}^{(3)}}_{n,:}\|_2\|\widehat{\boldsymbol{Y}}^{(3)}_{n,:}\|_2}\right),$$

which computes the angle between original and estimated fibers;

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

where μ_k^2 is the mean value of $\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}$. ERGAS represents the relative dimensional global error between the SRI and the estimate.

We also show the computational time for our implementation each algorithm, given by the `tic` and `toc` functions of MATLAB. As in [8], we run STEREO for 10 iterations.

5.2.1. Semi-real data

First, we compare the results for some semi-real data, available online at [19], with MSI and HSI generated artificially. The representative multilinear ranks are chosen to be $[40, 40, 6]$, $[30, 30, 16]$ and $[24, 24, 25]$, which correspond to different regions of identifiability; also the number of ground truth regions in the image is $E = 16$ hence the multilinear ranks correspond to different scenarios ($R_3 < E$, $R_3 = E$,$R_3 > E$). For STEREO, we choose ranks $F = 50$ and $F = 100$, and in HySure we use (as in [8]) 16 as the number of endmembers. Table 5.2.1 shows these metrics for the Indian Pines dataset mentioned above. In general, SCOTTand B-SCOTT(without partition into blocks) give comparable results with STEREO in the case $F = 50$ used in [8]. But, the case of STEREO $F = 100$ gives slightly better results than tensor rank $F = 50$, especially in terms of SNR. The B-SCOTTmethod appears to be fastest in this case (although we do not claim it, because the speed depends on the implementation), but needs higher multilinear ranks. Fast algorithms for Sylvester equations are used for SCOTTand STEREO.

Algorithm	R-SNR	CC	SAM	ERGAS	time
STEREO 50	26.93	0.89	2.25	1.03	8.06
STEREO 100	28.47	0.92	2.03	0.88	12.15
SCOTT(40,40,6)	26.32	0.89	2.34	1.07	3.27
SCOTT(30,30,16)	23.82	0.84	2.77	1.38	1.02
SCOTT(24,24,25)	24.65	0.87	2.56	1.23	0.55
B-SCOTT(40,40,6)	25.45	0.88	2.65	1.2	0.18
B-SCOTT(60,60,6)	26.06	0.89	2.57	1.14	0.1
B-SCOTT(100,100,6)	26.28	0.9	2.55	1.14	0.12
HySure	25.12	0.77	2.56	1.68	44.63

Table 1. Comparison of algorithms, Indian Pines dataset

The second dataset we consider is the Salinas A-scene, also acquired by the AVIRIS sensor; it is a portion of the bigger Salinas dataset. In this case, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{80\times84\times204}$, and $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{80\times84\times6}$, $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{20\times21\times204}$. In [8] $E = 6$ is quoted as number of endmembers (this value is chosed for HySure). The results are shown in Table 5.2.1. In this case, $R_3 = 6$ seems to be the best choice for SCOTT, which agrees with the number of materials. The SCOTTis again rather fast, except for the case where it shows a better performance than STEREO (with $F = 100$, which is the rank chosen in [8] for this dataset).

5.2.2. Real wood data

The third dataset is obtained from a wood plank. Instead of generating the MSI and HSI with specified dimensions, data cubes $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{640\times150\times224}$, $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{640\times150\times28}$ and $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{320\times75\times224}$ are acquired by a SPECIM FX17 hyperspectral camera, which spectral range is $[900\text{nm}, 1700\text{nm}]$. For more details on the camera, please refer to [20]. For this dataset,

Algorithm	R-SNR	CC	SAM	ERGAS	time
STEREO 50	33.63	0.98	0.91	1.39	4.03
STEREO 100	37.47	0.99	0.66	1.12	5.38
SCOTT(40,40,6)	33.59	0.99	0.77	1.18	2.19
SCOTT(14,14,15)	22.68	0.94	2.1	3.07	0.15
SCOTT(10,15,25)	23.17	0.96	2.1	2.59	0.14
SCOTT(30,30,6)	30.18	0.98	1.02	1.41	0.65
SCOTT$(58, 58, 6)$	37.55	0.99	0.61	1.03	14.7
HySure	29.05	0.97	1.39	3.28	14.53

Table 2. Comparison of algorithms, Salinas A-scene dataset

no groundtruth data is available. However, in most cases, the number of sources is considered very small ($E \leq 10$) [21]. We choose different values for R_3 for SCOTT, and for HySure, we choose $E = 8$. For STEREO we also try different ranks, because this dataset was not used in [8]. The results can be found in Table 5.2.2. We see that all the methods give comparable results, thus SCOTTcan be chosen, because it is especially fast for the case of small multilinear ranks. The comparison of the reconstruction of a selected spectral bands are shown in Fig. 7.

Algorithm	R-SNR	CC	SAM	ERGAS	time
STEREO 20	21.96	0.99	1.76	2.49	54.55
STEREO 30	21.93	0.99	1.75	2.5	49.79
STEREO 50	21.95	0.99	1.7	2.49	72.61
STEREO 100	21.94	0.99	1.6	2.49	62.75
SCOTT(50,50,8)	23.2	0.99	1.59	2.16	6.13
SCOTT(30,30,15)	23.12	0.99	1.7	2.18	3.44
HySure	21.73	0.99	1.91	2.56	186.28

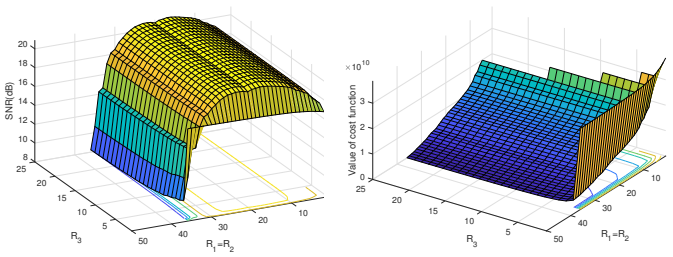
Table 3. Comparison of algorithms, wood dataset

5.2.3. Pansharpening problem

Here, we adress the pansharpening problem, which consists in fusion of a hyperspectral image and a panchromatic image (PAN) $\boldsymbol{\mathcal{Y}}_P$. PAN is obtained by averaging over the full spectral range of the groundtruth SRI, meaning that $\boldsymbol{P_M} \in \mathbb{R}^{1\times K}$ and $\boldsymbol{\mathcal{Y}}_P \in \mathbb{R}^{I\times J\times 1}$. In this case, the STEREO algorithm is not applicable, since its initialization[6] is based on the CPD of the MSI (which is a matrix in the case of PAN images). However, the coupled Tucker model is still recoverable when $R_1 \leq I_H$ and $R_2 \leq J_H$.

For the lest of this section, we consider the Indian Pines dataset. First, in Figure 6, we plot the SNR and cost for $R_1 = R_2$ in $[1 : 36]$ and R_3 in $[1 : 25]$ under the identifiability conditions. In Table 5.2.3, the metrics are shown for different multilinear ranks and for the HySure method. In this case as well, the previous conclusions on the metrics hold, showing that SCOTTis able to give a reasonable solution the pansharpening problem as well.

[6]We also tried different initializations for STEREO, including the one obtained based on our HOSVD solution.



SnR between SRI and estimate Cost value

Fig. 6. R-SNR and cost as a function of $R_1 = R_2$ and R_3

Algorithm	R-SNR	CC	SAM	ERGAS	time
SCOTT$(24, 24, 25)$	20.59	0.78	4.36	1.93	0.94
SCOTT$(30, 30, 16)$	18.48	0.7	5.37	2.48	1.08
SCOTT$(35, 35, 6)$	11.38	0.41	10.53	5.71	1.84
HySure	23.24	0.77	2.28	1.69	33.87

Table 4. Metrics for different algorithms, Indian Pines dataset

5.3. Endmember recovery

A correct recovery of spectral signatures is quite important for further processing of hyperspectral images, therefore we are going to see whether SCOTTis able to do that. We consider the Indian Pines dataser, where groundtruth data (see Fig. 9) 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 it with STEREO ($F = 16$).

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 endmembers 4,7,9 and 14, which are plotted in Figure 10. Note that materials 7 and 9 are scarce in the original SRI (resp. 28 and 20 pixels), whereas materials 4 and 14 are more abundant (resp. 237 and 1265 pixels).

In Figure 8 we plot relative errors of the reconstruction of spectra by different methods. As expected, for materials 7 and 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. In Figure 11, 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 estimates well the spectrum. For the scarce materials it is important to choose the rank, in particular $R_3 = 16$ and $R_3 = 25$ yield better reconstruction that $R_3 = 6$, and also STEREO with $F = 16$

5.4. Blind recovery of the SRI

In this subsection, we consider the blind recovery of the SRI. We compare the performance of SCUBA and B-SCOTTfor the Pavia University dataset. In this case, the SRI is acquired

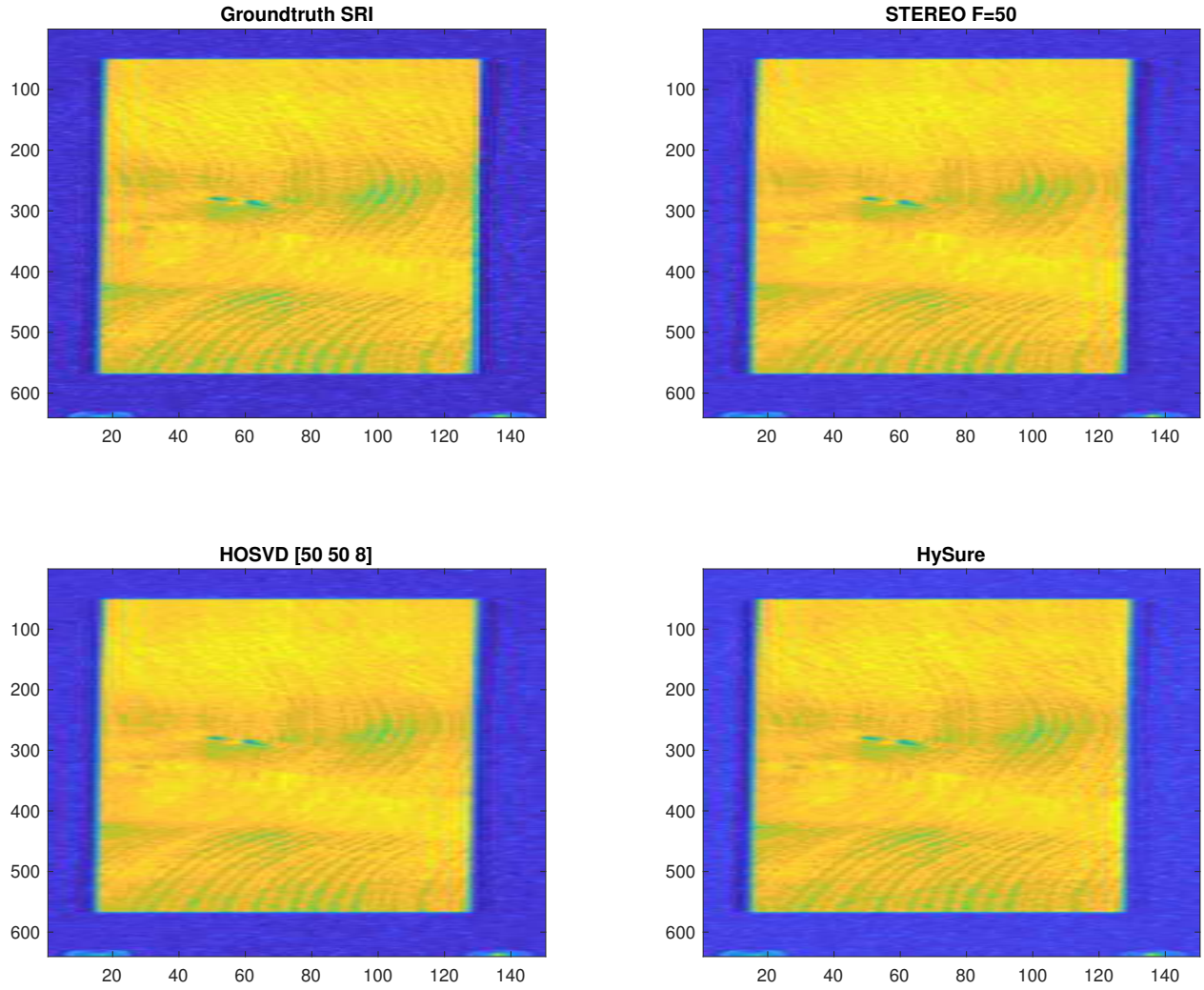


Fig. 7. Band 160 of SRI for different algorithms

from a ROSIS sensor such that $\mathcal{Y} \in \mathbb{R}^{608\times366\times103}$, $\mathcal{Y}_H \in \mathbb{R}^{152\times84\times103}$ and $\mathcal{Y} \in \mathbb{R}^{608\times366\times6}$. The results are given in Table 5.4, where for B-SCOTT"single" means that there is no partition into blocks, and $(4, 4)$ means that the images are partitioned into 16 blocks of the same size, according to a 4×4 grid. Reconstruction of one spectral band is shown in Fig. 12. Our results show that B-SCOTTcan approach the performance of SCUBA at a significantly lower computational cost.

Algorithm	R-SNR	CC	SAM	ERGAS	time
SCUBA 50	24.97	0.99	3.33	2.08	34.81
B-SCOTT$[100, 100, 4]$, single	16.6	0.94	6.03	4.62	0.47
B-SCOTT$[60, 60, 3]$ (4,4)	22.98	0.98	3.83	2.46	0.62
B-SCOTT$[120, 60, 4]$ (4,4)	24.21	0.99	3.44	2.09	0.8

Table 5. Comparison of blind algorithms, Pavia University dataset

6. CONCLUSION

In this paper, we proposed a novel coupled Tucker model for hyperspectral superresolution. We showed that the model is

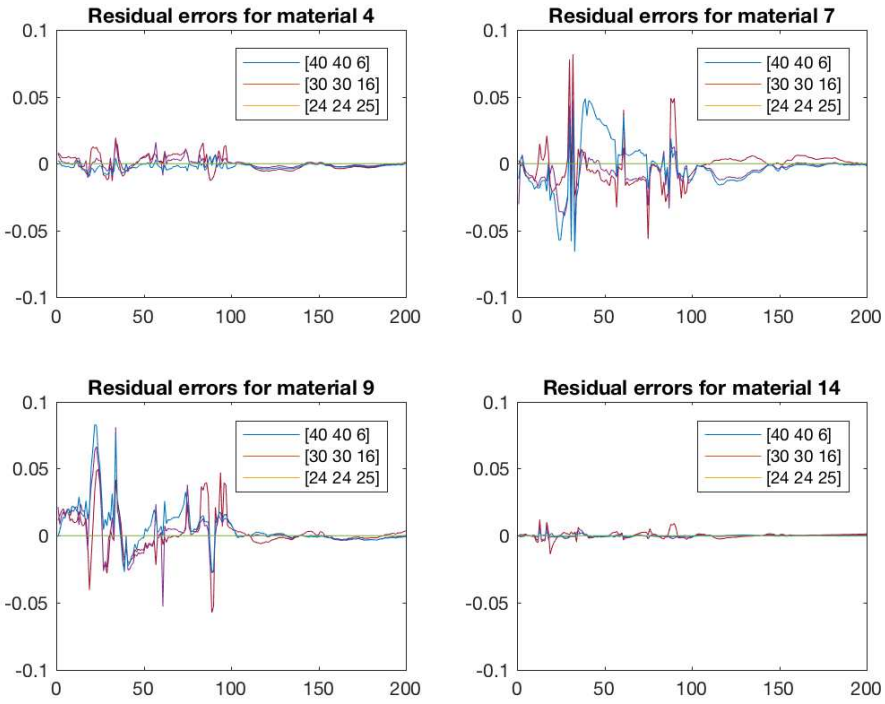


Fig. 8. Residual errors for the three considered ranks and four materials

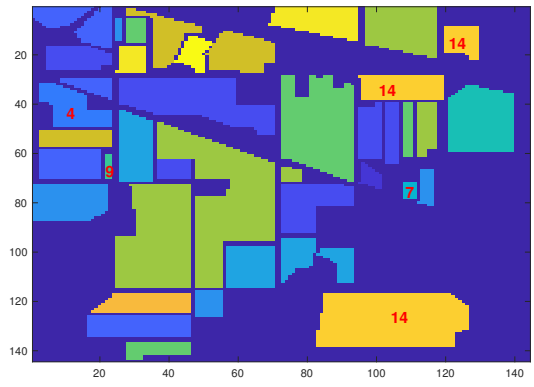


Fig. 9. Groundtruth image for Indian Pines dataset. Materials 4,7,9,14 are marked in red.

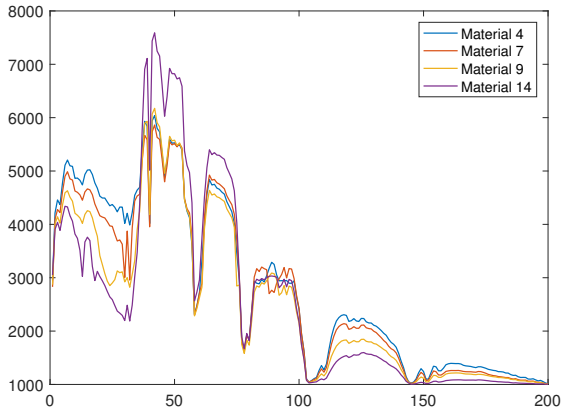


Fig. 10. Original spectral signature for materials 4,7,9 and 14

recoverable, that is, almost surely identifiable in a well chosen basis, and 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 problem) 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

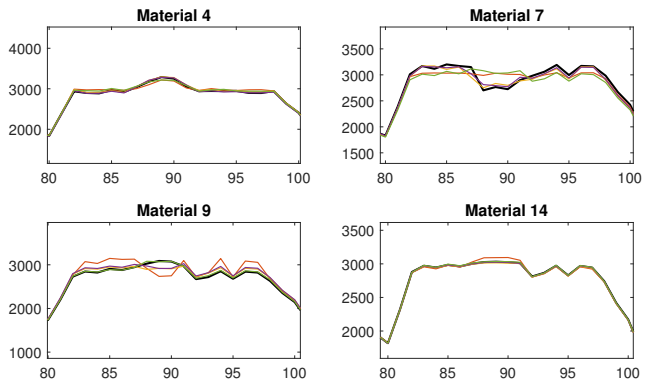


Fig. 11. Endmembers at spectral bins 80 to 100. Groundtruth (*black*), SCOTT$(40, 40, 6)$ (*red*), SCOTT$[30, 30, 16]$ (*yellow*), SCOTT$(24, 24, 25)$ (*purple*), STEREO $F = 100$ (*green*).

multilinear rank.

7. REFERENCES

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

[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 Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, Sept 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 Processing for Remote Sensing*, pp. 533–548, 2011.

[4] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, Feb 2012.

[5] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, 2015.

[6] M. Simoẽs, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Geoscience and Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, 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 Transactions on Signal Processing*, 2018, doi:10.1109/TSP.2018.2876362.

[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," Tech. Rep., 2018, arxiv:1807.10027.

[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 Conference on Image Processing (ICIP)*, Oct 2018, pp. 3318–3322.

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

[12] Tamara G. Kolda and Brett W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

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

[14] 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.

[15] G. Golub, S. Nash, and C. Van Loan, "A Hessenberg-Schur method for the problem AX+XB=C," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 909–913, 1979.

[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[17] Yang Qi, Pierre Comon, and Lek-Heng Lim, "Semialgebraic geometry of nonnegative tensor rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 37, no. 4, pp. 1556–1580, 2016.

[18] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, Available online.

[19] "Grupo de inteligencia computacional (gic)," `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`, accessed: 2018-10-23.

[20] "SPECIM FX17 homepage," http://www.specim.fi/downloads/Specim-FX17-datasheet-10102018.pdf, accessed: 2018-10-24.

[21] L. Nus, S. Miron, and D.Brie, "On-line blind unmixing for hyperspectral pushbroom imaging systems," *IEEE Statistical Signal Processing Workshop*, Jun 2018.

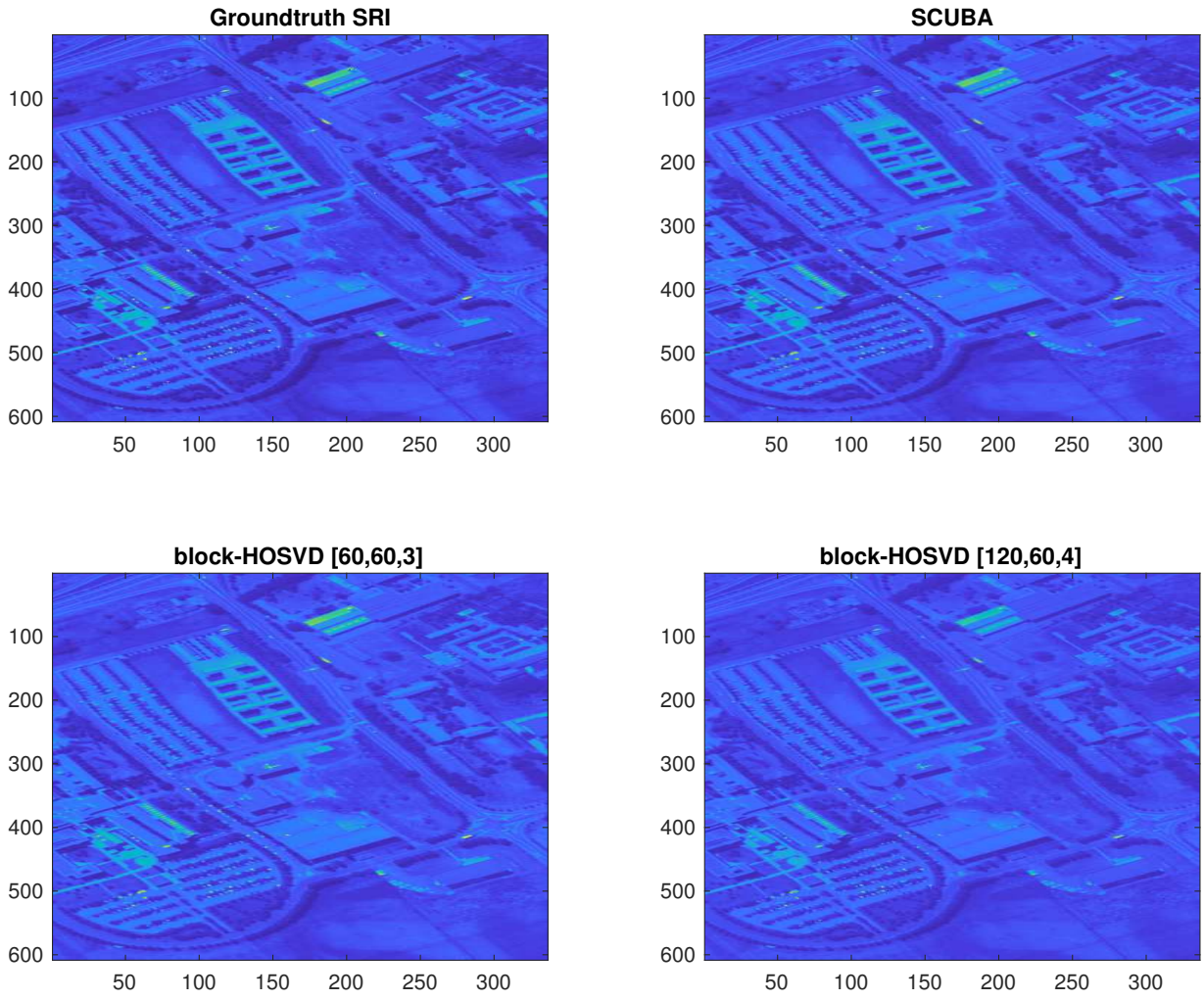


Fig. 12. Reconstruction of SRI at spectral band 44 for various algorithms