IMPROVED HYPERSPECTRAL SUPER-RESOLUTION BASED ON COUPLED TUCKER APPROXIMATION

C. Prévost, P. Chainais and R. Boyer

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

ABSTRACT

Hyperspectral super-resolution based on the coupled Tucker decomposition has been recently considered in the remote sensing community. Although competitive, the state-of-the-art approaches did not fully exploit the coupling information contained in hyperspectral and multispectral images of the same scene. In this paper, we propose a new algorithm that overcomes the limitations of the state-of-the-art. This method accounts for low-resolution information contained in the observations, by solving a set of least-squares problems. In addition, we provide exact recovery conditions for the super-resolution image in the noiseless case. Using a set of real datasets, we show that the proposed algorithm achieves good reconstruction with small complexity.

1. INTRODUCTION

Hyperspectral devices produce hyperspectral images (HSI) with high spectral resolution. However, the compromise between signal-to-noise ratio, spatial and spectral resolutions force the HSIs to have low spatial resolution [1]. On the other hand, multispectral images (MSI) have high spatial resolution, at the cost of a restricted number of spectral bands. The hyperspectral super-resolution (HSR) problem [2] was formulated to recover a super-resolution image (SRI) with both high spatial and high spectral resolutions from co-registered HSI and MSI of the same scene.

Early matrix-based approaches to HSR [3, 4, 5, 6] were based on the linear mixing model and performed a coupled low-rank factorization of the matricized HSI and MSI. More recently, tensor approaches were envisioned, motivated by the 3-dimensional structure of the observations, and possible uniqueness guarantees offered by tensor low-rank factorizations. The works of [7, 8] were the first to consider tensor-based HSR. Since then, various decompositions were considered, with possible constraints on the low-rank factors, see, e.g., [9, 10, 11, 12] and references therein.

The Tucker decomposition was steadily considered [13, 9, 12] for solving the HSR problem. It was proved in [9] that exact recovery of the SRI could be achieved in the noiseless case, even if the Tucker decomposition was not unique. Two algorithms based on the Singular Value Decomposition (SVD) were proposed in [9]. The first one was named Super-resolution based on COupled Tucker Tensor approximation (SCOTT). It recovered the Tucker factors based on several SVDs followed by solving a least-squares problem. The second one assumed that the spatial degradation between the HSI and SRI was unknown, therefore it was referred to as a blind version of SCOTT, namely BSCOTT (Blind-SCOTT). The Tucker model in [9] was later extended to account for localized changes in [12].

However, these algorithms have a major limitation: they did not fully exploit the coupling information between the low-resolution observations and the SRI. The blind algorithm BSCOTT only accounted for a portion of the coupling constraints between the observations. In contrast with BSCOTT, SCOTT was non-blind but was highly sub-optimal. Indeed, the SVD step only extracted high-resolution information from one single observation. Hence the low-resolution information available from the observation model was totally ignored. Moreover, since no iterations were performed, the solution to the least squares problem did not incorporate the coupling information to the Tucker factors.

In this paper, we propose a new Tucker-based algorithm that overcomes the technical limitations of SCOTT and BSCOTT. This new procedure fully accounts for the coupling constraints in the HSR observation model. We propose a simple way to incorporate additional information to the Tucker factors. The SVD is performed on both observations and is followed by solving a set of linear equations. We prove that exact recovery of the SRI is still achievable with the proposed approach, however for a restricted range of multilinear ranks. To circumvent this limitation, we propose a block-wise procedure that makes the rank conditions less restrictive. Our experiments on real datasets show that the proposed approach achieves good reconstruction performance with low complexity.

This paper is organized as follows. In Section 2 we introduce the coupled Tucker model and the proposed algorithm. Section 3 contains our recovery analysis for tensor reconstruction in the noiseless case. Finally, Section 4 contains the numerical experiments.

Notation. In this paper we mainly follow [14, 15] in what concerns the tensor notation. We use the symbol $\boxtimes$ for the Kronecker product, and $\odot$ for the Khatri-Rao product. We use $\mathrm{vec}\{\cdot\}$ for the standard column-major vectorization of a tensor or a matrix. By $\mathrm{tSVD}_R(\mathbf{X})$ we denote a matrix containing R leading right singular vectors of the matrix $\mathbf{X}$. Operation $\bullet_p$ denotes contraction on the pth index of a tensor; for instance, for a tensor $\boldsymbol{\mathcal{A}}$ and a matrix $\mathbf{M}$, $[\boldsymbol{\mathcal{A}} \bullet_1 \mathbf{M}]_{\ell jk} = \sum_\ell \mathcal{A}_{ijk} M_{i\ell}$. For a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$, its unfoldings are $\mathbf{Y}^{(1)} \in \mathbb{R}^{JK\times I}$, $\mathbf{Y}^{(2)} \in \mathbb{R}^{IK\times J}$ and $\mathbf{Y}^{(3)} \in \mathbb{R}^{IJ\times K}$. For a tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1\times R_2\times R_3}$ 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:

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

If, in addition, $R_1 = \mathrm{rank}(\mathbf{Y}^{(1)})$, $R_2 = \mathrm{rank}(\mathbf{Y}^{(2)})$ and $R_3 = \mathrm{rank}(\mathbf{Y}^{(3)})$, then the multilinear product is called Tucker decomposition of $\boldsymbol{\mathcal{Y}}$ and (R_1, R_2, R_3) are called the multilinear ranks.

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

2. HYPERSPECTRAL SUPER-RESOLUTION

2.1. Problem statement and observation model

We consider an MSI tensor $\mathcal{Y}_M \in \mathbb{R}^{I\times J\times K_M}$ and HSI tensor $\mathcal{Y}_H \in \mathbb{R}^{I_H\times J_H\times K}$ acquired from sensors (for instance, LANDSAT or QuickBird). The MSI has higher spatial resolution than the HSI ($I_H < I, J_H < J$), but it has some lower spectral resolution ($K_M < K$). The acquired MSI and HSI represent the same scene, and $\mathcal{Y}_M$ and $\mathcal{Y}_H$ are viewed as two degraded versions of a single SRI tensor $\mathcal{Y} \in \mathbb{R}^{I\times J\times K}$. The HSR problem consists in recovering $\mathcal{Y}$ from $\mathcal{Y}_M$ and $\mathcal{Y}_H$. As commonly adopted in the literature [7, 9, 10], we consider the following degradation model:

$$\begin{cases} \mathcal{Y}_H &= \mathcal{Y} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \mathcal{E}_H, \\ \mathcal{Y}_M &= \mathcal{Y} \bullet_3 \mathbf{P}_3 + \mathcal{E}_M, \end{cases} \tag{2}$$

where $\mathcal{E}_M$ and $\mathcal{E}_H$ are noise terms. The matrix $\mathbf{P}_3 \in \mathbb{R}^{K_M\times K}$ is the spectral degradation 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. This assumption is reasonable thanks to the commonly accepted Wald's protocol [16], that uses isotropic Gaussian blurring and downsampling for spatial degradation.

Similarly to [9], we utilize a Tucker-based coupled model to solve the reconstruction problem at hand. Assume that the SRI $\mathcal{Y}$ admits a Tucker decomposition with given (fixed) multilinear ranks $\boldsymbol{R} = (R_1, R_2, R_3)$ as

$$\mathcal{Y} = [\![\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!], \tag{3}$$

where $\mathbf{U} \in \mathbb{R}^{I\times R_1}$, $\mathbf{V} \in \mathbb{R}^{J\times R_2}$ and $\mathbf{C} \in \mathbb{R}^{K\times R_3}$ 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, the degradation model (2) becomes

$$\begin{cases} \mathcal{Y}_H &= [\![\mathcal{G}; \mathbf{P}_1\mathbf{U}, \mathbf{P}_2\mathbf{V}, \mathbf{W}]\!] + \mathcal{E}_H, \\ \mathcal{Y}_M &= [\![\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{P}_3\mathbf{W}]\!] + \mathcal{E}_M. \end{cases} \tag{4}$$

The aim of HSR is to recover the factor matrices $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ and the core tensor $\mathcal{G}$ from the coupled Tucker model (4).

2.2. Tucker-based HSR: state-of-the-art and its limitations

In [9], an algorithm was proposed to solve the Tucker-based HSR problem. This algorithm, called Super-resolution based on COupled Tucker Tensor approximation (SCOTT), consisted in three simple steps. First, the factor matrices $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$, $\widehat{\mathbf{W}}$ were recovered as the dominant right singular vectors of the unfoldings $\mathbf{Y}_M^{(1)}$, $\mathbf{Y}_M^{(2)}$ and $\mathbf{Y}_H^{(3)}$, respectively. Then, the core tensor $\widehat{\mathcal{G}}$ was reconstructed as:

$$\begin{aligned} \underset{\mathcal{G}}{\operatorname{argmin}}\ f_T(\mathcal{G}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}) = \|\mathcal{Y}_H - [\![\mathcal{G}; \mathbf{P}_1\widehat{\mathbf{U}}, \mathbf{P}_2\widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!]\|_F^2 \\ + \lambda\|\mathcal{Y}_M - [\![\mathcal{G}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \mathbf{P}_3\widehat{\mathbf{W}}]\!]\|_F^2. \end{aligned} \tag{5}$$

This step consists in solving the following least-squares problem:

$$\underbrace{\begin{bmatrix} \widehat{\mathbf{W}} \boxtimes \mathbf{P}_2\widehat{\mathbf{V}} \boxtimes \mathbf{P}_1\widehat{\mathbf{U}} \\ \sqrt{\lambda}\mathbf{P}_M\widehat{\mathbf{W}} \boxtimes \widehat{\mathbf{V}} \boxtimes \widehat{\mathbf{U}} \end{bmatrix}}_{\mathbf{X}} \operatorname{vec}\{\widehat{\mathcal{G}}\} \approx \underbrace{\begin{bmatrix} \operatorname{vec}\{\mathcal{Y}_H\} \\ \sqrt{\lambda}\operatorname{vec}\{\mathcal{Y}_M\} \end{bmatrix}}_{\boldsymbol{z}},$$

that can be solved through normal equations of the form

$$\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right) \operatorname{vec}\{\widehat{\mathcal{G}}\} = \mathbf{X}^\mathsf{T}\boldsymbol{z}. \tag{6}$$

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

$$\begin{aligned} \mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{I}_{R_3} \boxtimes \left(\widehat{\mathbf{V}}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\mathbf{P}_2\widehat{\mathbf{V}}\right) \boxtimes \left(\widehat{\mathbf{U}}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\mathbf{P}_1\widehat{\mathbf{U}}\right) \\ + \lambda\left(\widehat{\mathbf{W}}^\mathsf{T}\mathbf{P}_M^\mathsf{T}\mathbf{P}_M\widehat{\mathbf{W}}\right) \boxtimes \mathbf{I}_{R_1R_2}, \end{aligned} \tag{7}$$

and the vector on the right-hand side is

$$\begin{aligned} \mathbf{X}^\mathsf{T}\boldsymbol{z} = \operatorname{vec}\{[\![\mathcal{Y}_H; \widehat{\mathbf{U}}^\mathsf{T}\mathbf{P}_1^\mathsf{T}, \widehat{\mathbf{V}}^\mathsf{T}\mathbf{P}_2^\mathsf{T}, \widehat{\mathbf{W}}^\mathsf{T}]\!]\} \\ + \lambda \operatorname{vec}\{[\![\mathcal{Y}_M; \widehat{\mathbf{U}}^\mathsf{T}, \widehat{\mathbf{V}}^\mathsf{T}, \widehat{\mathbf{W}}^\mathsf{T}\mathbf{P}_M^\mathsf{T}]\!]\}. \end{aligned} \tag{8}$$

This step could be seen as solving a (generalized) Sylvester equation, for which fast solvers were used, see e.g., [17] and references therein. Finally, the low-rank approximation of the SRI was reconstructed using (3). In [9], experiments on real images showed the capabilities of SCOTT to solve the HSR problem with a low computational complexity.

However, SCOTT suffered from major limitations. First, the estimation of the factor matrices was sub-optimal. For instance, $\widehat{\mathbf{U}}$ was recovered only from $\mathcal{Y}_M$, hence the low-resolution spatial information contained in $\mathcal{Y}_H$ was totally ignored. Recovery of $\widehat{\mathbf{V}}$ and $\widehat{\mathbf{W}}$ suffered from the same drawbacks. In other words, the second step of SCOTT did not incorporate at all the additional low-resolution information. As a result, SCOTT did not consider all the available information from (4).

2.3. Proposed approach

In this paper, we propose an improved Tucker-based algorithm that overcomes the limitations of SCOTT. In contrast with SCOTT, the estimation strategy for $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ accounts for the low-resolution information from the observations. This strategy was considered in [8] for tensor-matrix couplings. In BSCOTT[1] [9], it was for estimation of $\mathbf{W}$ only.

One specificity of model (2) is that high-resolution information in one given dimension is always available in at least one observation. The full spectral information can be obtained from $\mathbf{Y}_H^{(3)}$, while the full spatial information is contained in $\mathbf{Y}_M^{(1)}$ and $\mathbf{Y}_M^{(2)}$. Therefore, the recovery of the factor matrices $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ underlying the SRI will be driven by the high-resolution observations. The low-resolution information will also be taken into account. In the following, we will explain how to obtain $\mathbf{W}$, and the factor matrices $\mathbf{U}$ and $\mathbf{V}$ can be obtained similarly.

First, let us consider the third unfolding of the HSI, which contains the full spectral information. This matrix admits a SVD as

$$\mathbf{Y}_H^{(3)} = \mathbf{L}_H\boldsymbol{\Sigma}_H\mathbf{W}_H^\mathsf{T}, \tag{9}$$

where $\mathbf{W}_H \in \mathbb{R}^{K\times R_3}$ contains the dominant right singular vectors of $\mathbf{Y}_H^{(3)}$, and gives an orthogonal basis for the fiberspace of $\mathcal{Y}_H$. Equation 9 offers a solution for recovery of the full spectral information, and boils down to solving least-squares problems [18].

Note that, following (2), the HSI $\mathcal{Y}_H$ and the SRI $\mathcal{Y}$ share the same fiberspace. Hence we can express the third unfolding of $\mathcal{Y}$ as

$$\mathbf{Y}^{(3)} = \mathbf{L}\boldsymbol{\Sigma}\mathbf{W}_H^\mathsf{T}. \tag{10}$$

By unfolding (3) in the third dimension, we have

$$\mathbf{Y}^{(3)} = \left[(\mathbf{V} \boxtimes \mathbf{U})\mathbf{G}^{(3)}\right] \mathbf{W}^\mathsf{T},$$

[1] In [8] and BSCOTT, the spatial degradation was considered to be unknown, hence the reconstruction problem different.

which means that $\mathbf{W}$ lives in a low-dimensional subspace defined by $\mathbf{W}_H$ [19]. In other words, it holds that $\mathbf{W}_H\mathbf{W}_H^\mathsf{T}\mathbf{W}^\mathsf{T} = \mathbf{W}^\mathsf{T}$.

Assuming that the subspace containing the spectral information is of dimension R_3, the equation (10) allows $\mathcal{Y}$ to be compressed without loss of spectral information to a tensor $\mathcal{Z} \in \mathbb{R}^{I\times J\times R_3}$ such as

$$\mathcal{Z} = \mathcal{Y} \bullet_3 \mathbf{W}_H^\mathsf{T}.$$

Note that we can recover the SRI as $\mathcal{Y} = \mathcal{Z} \bullet_3 \mathbf{W}_H$. According to (3), the third unfolding of $\mathcal{Z}$ can be written as

$$\mathbf{Z}^{(3)} = \left[(\mathbf{V} \boxtimes \mathbf{U})\mathbf{G}^{(3)}\right] \mathbf{W}^\mathsf{T}\mathbf{W}_H = \left[(\mathbf{V} \boxtimes \mathbf{U})\mathbf{G}^{(3)}\right] \left(\mathbf{W}_H^\mathsf{T}\mathbf{W}\right)^\mathsf{T}.$$

Let us now consider the low-resolution information contained in the MSI $\mathcal{Y}_M$. From (2) in the noiseless case, we can write

$$\mathcal{Y}_M = \mathcal{Y} \bullet_3 \mathbf{P}_3 = (\mathcal{Z} \bullet_3 \mathbf{W}_H) \bullet_3 \mathbf{P}_3 = \mathcal{Z} \bullet_3 \mathbf{P}_3\mathbf{W}_H.$$

Therefore, the third unfolding of the MSI can be expressed using the SVD as

$$\mathbf{Y}_M^{(3)} = \mathbf{L}_M\mathbf{\Sigma}_M \left(\mathbf{P}_3\mathbf{W}_H\mathbf{W}_H^\mathsf{T}\mathbf{W}\right)^\mathsf{T} = \mathbf{L}_M\mathbf{\Sigma}_M\mathbf{W}_M^\mathsf{T}.$$

where the matrix $\mathbf{W}_M = \mathbf{P}_3\mathbf{W}_H\mathbf{W}_H^\mathsf{T}\mathbf{W}$ corresponds to the orthogonal projection of $\mathbf{W}$ on the fiberspace of $\mathcal{Y}$, followed by spectral degradation. This matrix is computed from the SVD of $\mathbf{Y}_M^{(3)}$.

Since $\mathbf{W}$ lives in the span of $\mathbf{W}_H$, and under the assumption that $R_3 \leq K_M$, the factor matrix $\mathbf{W}_H^\mathsf{T}\mathbf{W}$ is recovered by solving the overdetermined $\mathbf{W}_M = \mathbf{P}_3\mathbf{W}_H\left(\mathbf{W}_H^\mathsf{T}\mathbf{W}\right)$. The matrix $\mathbf{W}$ will be obtained up to a term belonging to the span of $\mathbf{W}_H$. Since the SVD provides the solution that minimizes the distance between the span of $\mathbf{W}$ and that of $\mathbf{W}_H$, we finally obtain

$$\mathbf{W} = \mathbf{W}_H\left(\mathbf{P}_3\mathbf{W}_H\right)^\dagger\mathbf{W}_M.$$

The proposed approach is summarized in Algorithm 1.

Algorithm 1: Proposed approach

input : $\mathcal{Y}_M \in \mathbb{R}^{I\times J\times K_M}$, $\mathcal{Y}_H \in \mathbb{R}^{I_H\times J_H\times K}$, R_1, R_2, R_3

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

1. $\mathbf{U}_M \leftarrow \text{tSVD}_{R_1}\left(\mathbf{Y}_M^{(1)}\right)$, $\mathbf{U}_H \leftarrow \text{tSVD}_{R_1}\left(\mathbf{Y}_H^{(1)}\right)$, $\widehat{\mathbf{U}} \leftarrow \mathbf{U}_M(\mathbf{P}_1\mathbf{U}_M)^\dagger\mathbf{U}_H$;
2. $\mathbf{V}_M \leftarrow \text{tSVD}_{R_2}\left(\mathbf{Y}_M^{(2)}\right)$, $\mathbf{V}_H \leftarrow \text{tSVD}_{R_2}\left(\mathbf{Y}_H^{(2)}\right)$, $\widehat{\mathbf{V}} \leftarrow \mathbf{V}_M(\mathbf{P}_2\mathbf{V}_M)^\dagger\mathbf{V}_H$;
3. $\mathbf{W}_M \leftarrow \text{tSVD}_{R_3}\left(\mathbf{Y}_M^{(3)}\right)$, $\mathbf{W}_H \leftarrow \text{tSVD}_{R_3}\left(\mathbf{Y}_H^{(3)}\right)$, $\widehat{\mathbf{W}} \leftarrow \mathbf{W}_H(\mathbf{P}_3\mathbf{W}_H)^\dagger\mathbf{W}_M$;
4. $\widehat{\mathcal{G}} \leftarrow \underset{\mathcal{G}}{\text{argmin}}\ f_T(\mathcal{G}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}})$;
5. $\widehat{\mathcal{Y}} = [\![\widehat{\mathcal{G}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!]$.

The ranks (R_1, R_2, R_3) cannot exceed (I_H, J_H, K_M), because the truncated SVD is performed on the degraded unfoldings $\mathbf{Y}_H^{(1)}$, $\mathbf{Y}_H^{(2)}$, $\mathbf{Y}_M^{(3)}$. As suggested in [8], we circumvent this limitation by applying Algorithm 1 to corresponding non-overlapping subblocks of the MSI and the HSI. This is based on the hypothesis that smaller blocks in the observations are more likely to contain a small number of materials and spatial features. Thus, $\mathcal{Y}_H$ and $\mathcal{Y}_M$ are divided into corresponding $L \times L$ subblocks of spatial dimensions $\frac{I_H}{L} \times \frac{J_H}{L}$ and $\frac{I}{L} \times \frac{J}{L}$, respectively. This strategy is summarized in Algorithm 2.

Algorithm 2: Block version of Algorithm 1

input : $\mathcal{Y}_M \in \mathbb{R}^{I\times J\times K_M}$, $\mathcal{Y}_H \in \mathbb{R}^{I_H\times J_H\times K}$, $\mathbf{R}$

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

Split $\mathcal{Y}_H$ and $\mathcal{Y}_M$ into corresponding subblocks;

for $\ell = 1, \ldots, L$ **do**

  Apply Algorithm 1 to the pair of blocks of $\mathcal{Y}_H$ and $\mathcal{Y}_M$, and store the result into the corresponding block of $\widehat{\mathcal{Y}}$.

end

The total computational complexity of Algorithm 1 is

- $\mathcal{O}(\min(R_1, R_2, R_3)(IJK_M + I_HJ_HK))$ flops for the truncated SVDs;
- $\mathcal{O}(\min(R_3^3 + (R_1R_2)^3; R_1^3 + (R_2R_3)^3))$ flops for solving the Sylvester equation in Step 4.

The computational complexity of Algorithm 1 is dominated by the cost of the truncated SVD. However, it can be smaller than that of SCOTT, which required $\mathcal{O}(\min(R_1, R_2)IJK_M + R_3I_HJ_HK)$ flops for recovery of the factor matrices.

3. RECOVERABILITY OF THE COUPLED MODEL

In this subsection, we give uniqueness results for the SRI tensor recovery by the approach of Algorithm 2.

3.1. Deterministic recovery

We begin with some deterministic results.

Theorem 3.1. *Let a Tucker decomposition of $\mathcal{Y}$ with multilinear ranks (R_1, R_2, R_3) be*

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

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

$$\text{rank}\left(\mathbf{Y}_M^{(1)}\right) = R_1,\ \text{rank}\left(\mathbf{Y}_M^{(2)}\right) = R_2,\ \text{rank}\left(\mathbf{Y}_H^{(3)}\right) = R_3, \quad (11)$$

and if

$$\text{rank}\left(\mathbf{P}_1\mathbf{U}\right) = R_1,\ \text{rank}\left(\mathbf{P}_2\mathbf{V}\right) = R_2,\ \text{rank}\left(\mathbf{P}_3\mathbf{W}\right) = R_3, \quad (12)$$

then Algorithm 1 recovers $\mathcal{Y}$ correctly. That is, there exists only one $\widehat{\mathcal{Y}}$ with multilinear ranks (R_1, R_2, R_3) such that $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ and $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_3 = \mathcal{Y}_M$.

Conversely, if

$$\text{rank}(\mathbf{P}_1\mathbf{U})\text{rank}(\mathbf{P}_2\mathbf{U}) < R_1R_2 \text{ and } \text{rank}(\mathbf{P}_3\mathbf{W}) < R_3 \quad (13)$$

then there exists infinitely many $\widehat{\mathcal{Y}}$ of the form

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

such that $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ and $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_3 = \mathcal{Y}_M$.

Proof. First, let us note that the conditions (11)–(12) in Theorem 3.1 allows for recovery of the factor matrices up to a change of basis. In the following, we will prove it for $\mathbf{U}$.
Indeed, by (4) we have $\boldsymbol{\mathcal{Y}}_H = [\![\boldsymbol{\mathcal{G}}; \mathbf{P}_1\mathbf{U}, \mathbf{P}_2\mathbf{V}, \mathbf{W}]\!]$. Since rank $(\mathbf{P}_1\mathbf{U}) = R_1$ and rank $(\mathbf{P}_2\mathbf{V}) = R_2$, then rank $\left(\mathbf{Y}_H^{(1)}\right) = R_1$ and rank $\left(\mathbf{Y}_H^{(2)}\right) = R_2$. Moreover, by (11) rank $\left(\mathbf{Y}_H^{(3)}\right) = R_3$. Therefore the multilinear rank of $\boldsymbol{\mathcal{Y}}_H$ is (R_1, R_2, R_3), which is equal to that of $\boldsymbol{\mathcal{Y}}$. Moreover,

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Y}}_H \underset{1}{\bullet} \left(\mathbf{U}(\mathbf{P}_1\mathbf{U})^\dagger\right) \underset{2}{\bullet} \left(\mathbf{V}(\mathbf{P}_2\mathbf{V})^\dagger\right).$$

Due to rank $(\mathbf{P}_1\mathbf{U}) = R_1$, Algorithm 1 recovers $\mathbf{U}$ up to a change of basis, *i.e.*, $\mathbf{Z} = \mathbf{U}\mathbf{O}$, where $\mathbf{O} \in \mathbb{R}^{R_1 \times R_1}$ is an orthogonal matrix. Finally, we have

$$\mathbf{U}\mathbf{O}(\mathbf{P}_1\mathbf{U}\mathbf{O})^\dagger = \mathbf{U}(\mathbf{P}_1\mathbf{U})^\dagger,$$

which completes the proof for $\mathbf{U}$. The proof for $\mathbf{V}$ and $\mathbf{W}$ can be obtained likewise.

Now, let us prove how (13) implies the non-uniqueness of $\widehat{\boldsymbol{\mathcal{Y}}}$. It should be noted that the core tensor $\widehat{\boldsymbol{\mathcal{G}}}$ can be obtained as a solution to (5) through normal equations of the form (6). where $\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{I}_{R_3} \boxtimes \mathbf{A} + \mathbf{B} \boxtimes \mathbf{I}_{R_1R_2}$. By [20, Theorem 13.16], the singular values of $\mathbf{X}^\mathsf{T}\mathbf{X}$ are all sums of the pairs of eigenvalues of

$$\mathbf{A} = \left(\widehat{\mathbf{V}}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\mathbf{P}_2\widehat{\mathbf{V}}\right) \boxtimes \left(\widehat{\mathbf{U}}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\mathbf{P}_1\widehat{\mathbf{U}}\right), \ \mathbf{B} = \lambda\widehat{\mathbf{W}}^\mathsf{T}\mathbf{P}_3^\mathsf{T}\mathbf{P}_3\widehat{\mathbf{W}}. \tag{14}$$

We also assume without loss of generality that $\mathbf{U}, \mathbf{V}, \mathbf{W}$ have orthonormal columns. Assume that rank$(\mathbf{P}_1\mathbf{U})$rank$(\mathbf{P}_2\mathbf{U}) < R_1R_2$ and rank$(\mathbf{P}_3\mathbf{W}) < R_3$. If we set $\widehat{\mathbf{U}} = \mathbf{U}$, $\widehat{\mathbf{V}} = \mathbf{V}$, $\widehat{\mathbf{W}} = \mathbf{W}$, then rank$(\mathbf{A}) < R_1R_2$, rank$(\mathbf{B}) < R_3$ and rank$(\mathbf{X}^\mathsf{T}\mathbf{X}) < R_1R_2R_3$. Therefore the system (6) is underdetermined, and there is an infinite number of solutions $\widehat{\boldsymbol{\mathcal{G}}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$. Note that if we define $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$, then it is an admissible solution, i.e., $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 = \boldsymbol{\mathcal{Y}}_M$ and $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \boldsymbol{\mathcal{Y}}_H$. Due to orthogonality of the bases, $\|\widehat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}\|_F = \|\widehat{\boldsymbol{\mathcal{G}}} - \boldsymbol{\mathcal{G}}\|_F$, which can be made arbitrary large due to non-uniqueness of the solution to (6). Hence the proof is complete.

□

3.2. Generic recovery

Theorem 3.2. *Assume that $\mathbf{P_1} \in \mathbb{R}^{I_H \times I}$, $\mathbf{P_2} \in \mathbb{R}^{J_H \times J}$, and $\mathbf{P_3} \in \mathbb{R}^{K_M \times K}$ are fixed full row-rank matrices. Let*

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

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$, $R_1 \leq I$, $R_2 \leq J$, $R_3 \leq K$, and $\mathbf{U} \in \mathbb{R}^{I \times R_1}$, $\mathbf{V} \in \mathbb{R}^{J \times R_2}$, $\mathbf{W} \in \mathbb{R}^{K \times R_3}$ are random matrices, distributed according to an absolutely continuous probability distribution. We also assume that $\boldsymbol{\mathcal{E}}_M, \boldsymbol{\mathcal{E}}_H = \mathbf{0}$ in (2)*.*

If $(R_1, R_2, R_3) \leq (I_H, J_H, K_M)$, and

$$R_1 \leq R_2R_3, \ R_2 \leq R_1R_3 \text{ and } R_3 \leq R_1R_2, \tag{15}$$

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

Proof. First, without loss of generality, we can replace $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ with the following of same size (see [9, Section V.B]):

$$\widetilde{\mathbf{P}}_1 = \begin{bmatrix} \mathbf{I}_{I_H} \\ \mathbf{0} \end{bmatrix}^\mathsf{T}, \widetilde{\mathbf{P}}_2 = \begin{bmatrix} \mathbf{I}_{J_H} \\ \mathbf{0} \end{bmatrix}^\mathsf{T}, \widetilde{\mathbf{P}}_3 = \begin{bmatrix} \mathbf{I}_{K_M} \\ \mathbf{0} \end{bmatrix}^\mathsf{T}. \tag{16}$$

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

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

Next, we are going to show how (15) imply (11). We will prove it only for the first condition (the others are analogous). Note that the first unfolding can be written as

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

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, $\mathbf{Y}_H^{(1)}$ will be generically of rank R_1 if we can provide just a single example of $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, $\boldsymbol{\mathcal{G}}$, achieving the condition 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. □

Remark 3.3. *The above conditions for exact recovery of the SRI were formulated in the noiseless case. This assumption, although rather strong, can be justified by the fact that real hyperspectral and multispectral images tend to have a high signal-to-noise ratio. In practice, the rank conditions given in Theorem 3.2 guarantee exact reconstruction under some* small *additive noise, hence the multilinear ranks must be chosen accordingly. This matter will be discussed in Section 4.1.*

4. EXPERIMENTS

All simulations were run on a MacBook Pro with 2.4 GHz Intel Core i5 and 8GB RAM. For basic tensor operations we used TensorLab 3.0 [21].

The reconstruction performance was evaluated using various metrics used in [22], including Reconstruction Signal-to-noise Ratio (R-SNR), Cross-Correlation (CC), Spectral Angle Mapper (SAM) and Relative dimensionless Global Error in Synthesis (ERGAS). The computation time was evaluated using the `tic` and `toc` functions of MATLAB.

The spectral bands corresponding to water absorption were removed. The HSI $\boldsymbol{\mathcal{Y}}_H$ was obtained by spatial degradation of the true SRI by $\mathbf{P}_1$ and $\mathbf{P}_2$ and the MSI $\boldsymbol{\mathcal{Y}}_M$ was obtained by spectral degradation by $\mathbf{P}_3$. White Gaussian noise was added to the observations to yield 30dB SNR. The spectral bands of $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$ were normalized and the true SRI was denoised [23]. The degradation matrices $\mathbf{P_1}, \mathbf{P_2}$ were generated following the commonly used Wald's protocol [16] with a downsampling ratio d and a Gaussian kernel of size $q = 9$. The matrix $\mathbf{P}_3$ contained the spectral response functions of the Sentinel-2 instrument[2]. We compared our algorithm to matrix-based approaches, including HySure [5], CNMF [3], GLP-HS [24] and FUSE [25]. We chose the ranks and regularization parameters according to the original works. We also considered tensor methods, namely STEREO [7] for CP decomposition, SCOTT and BSCOTT [9] for Tucker and CB-STAR for block-term decomposition [12]. For these algorithms, we chose the ranks according to [12] whenever possible.

[2]Available for download at `https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library`.

4.1. Choice of the ranks

We considered a portion of the Lockwood dataset with $\mathcal{Y} \in \mathbb{R}^{88\times88\times173}$. The HSI $\mathcal{Y}$ was produced by degradation with a downsampling ratio $d = 2$. We studied the impact of the multilinear ranks on the performance of Algorithm 2. We considered two scenarios: the first one splits the observations into $[2, 2]$ blocks and the second one considers $[4, 4]$ corresponding blocks. For each scenario, we computed the R-SNR as a function of $R_1 = R_2$ and R_3 for ranks satisfying Theorem 3.2. The results were displayed in Figure 1.

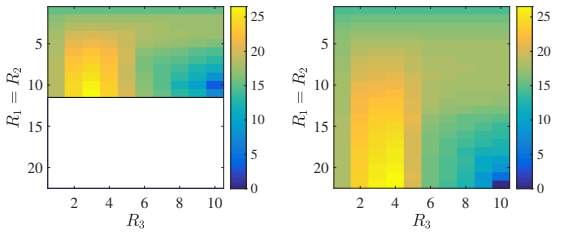

Fig. 1. R-SNR as a function of $R_1 = R_2$ and R_3, $[4, 4]$ (left) and $[2, 2]$-block pattern (right), Lockwood.

In both cases, good performance could be achieved even if the rank conditions were very restrictive. The best performance was obtained for small R_3, respectively $R_3 = 3$ and $R_3 = 4$, and large $R_1 = R_2$ inside the exact recovery region. A drop of performance could be observed for large R_3. To explain this phenomenon, we studied the singular values of the third unfolding of the HSI, namely $\mathbf{Y}_H^{(3)}$ for the $[2, 2]$- and $[4, 4]$-block patterns. We averaged the singular values of the unfoldings of the subblocks of the HSI. The first 30 singular values were displayed in Figure 2.

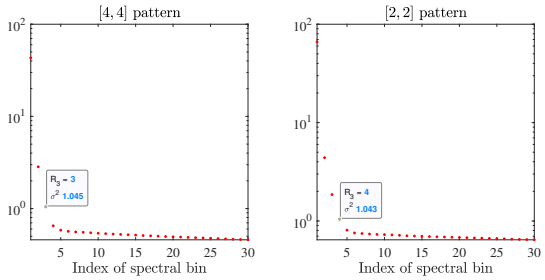

Fig. 2. Average singular values of subblocks of $\mathbf{Y}_H^{(3)}$, $[4, 4]$ (left) and $[2, 2]$-block pattern (right), Lockwood.

The elbows in the curves were coherent with the best R_3 obtained from Figure 1. Selecting higher R_3 would preserve some of the least significant singular values, leading the reconstruction to be corrupted by noise. In following experiments, the multilinear ranks were chosen according to Figure 1 so that they maximized the R-SNR, and likewise for other datasets. The ranks are summarized in Table 1.

4.2. HSI-MSI fusion

In this subsection, we assessed the performance of our approach for fusion of an HSI and an MSI. We first considered the Lockwood dataset introduced in Section 4.1. We took $[4, 4]$ and $[2, 2]$-block patterns and we chose the multilinear ranks $(11, 11, 3)$ and $(22, 22, 4)$, respectively. We also ran BSCOTT with $[4, 4]$ blocks and ranks $(11, 11, 3)$, for comparison. The reconstruction metrics and computation time were shown in Table 2, and the two best results of each columns were displayed in bold. The numbers between brackets for BSCOTT and Algorithm 2 denoted the number of blocks. In Figure 3, we showed false color plots of the true and reconstructed SRI by various algorithms, including Algorithm 2.

Dataset	SCOTT	STEREO	CB-STAR
Lockwood	$(60, 60, 5)$	$F = 50$	(70,70,5) (40,40,3)
Isabella Lake	$(60, 60, 5)$	$F = 50$	(50,50,5) (40,40,3)
Pavia University	$(40, 40, 1)$	n/a	(50,50,1) (40,40,1)

Table 1. Tensor-based ranks used in the experiments.

Algorithm	R-SNR	CC	SAM	ERGAS	Time
Best	∞	1	0	0	0
SCOTT	22.136	0.9464	3.28202	14.409	2.364
BSCOTT $[4, 4]$	26.705	0.9754	**1.9568**	10.901	0.1765
Alg. 3 $[4, 4]$	26.259	0.9745	1.9779	11.096	**0.1384**
Alg. 3 $[2, 2]$	26.544	**0.9773**	**1.9626**	**10.14**	**0.1452**
STEREO	**27.604**	0.9773	2.1199	19.494	2.255
CB-STAR	**27.626**	**0.9805**	1.8865	11.573	51.58
CNMF	22.663	0.9641	2.3962	14.095	4.605
GLP-HS	20.665	0.9387	2.5379	15.084	6.892
HySure	22.143	0.9470	3.4563	22.076	13.000
FUSE	23.006	0.9544	2.8842	**9.6755**	0.4874

Table 2. Reconstruction metrics, Lockwood.

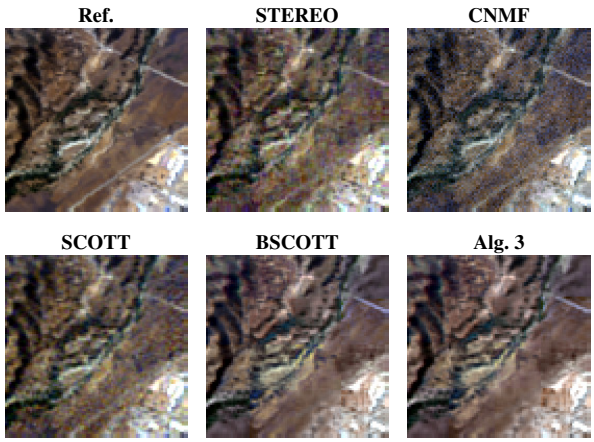

Fig. 3. False color plots of reconstructed SRI, Lockwood.

The second dataset was Isabella Lake with $\mathcal{Y} \in \mathbb{R}^{88\times120\times173}$. We used a downsampling ratio $d = 2$ for the HSI. For Algorithm 2 and BSCOTT, we considered ranks $(11, 11, 5)$ and $(22, 22, 5)$ for the $[4, 4]$ and $[2, 2]$ splitting scenarios, respectively. The results were available in Table 3 and Figure 4.

Algorithm	R-SNR	CC	SAM	ERGAS	Time
Best	∞	1	0	0	0
SCOTT	25.360	0.9842	3.5027	9.3365	2.270
BSCOTT $[4,4]$	**27.663**	0.9892	**2.4672**	**6.6404**	0.1707
Alg. 3 $[4,4]$	27.319	0.9891	**2.4733**	**6.6369**	**0.1523**
Alg. 3 $[2,2]$	27.374	**0.9895**	2.7336	6.9333	**0.1589**
STEREO	**29.119**	**0.9902**	2.7780	9.8351	2.902
CB-STAR	26.788	0.9881	2.9085	7.7640	15.07
CNMF	24.975	0.9882	2.5073	8.3610	5.737
GLP-HS	21.483	0.9742	3.2290	9.1432	9.845
HySure	22.023	0.9706	5.4772	12.8442	15.81
FUSE	24.505	0.9846	3.0626	7.6932	0.7477

Table 3. Reconstruction metrics, Isabella Lake.

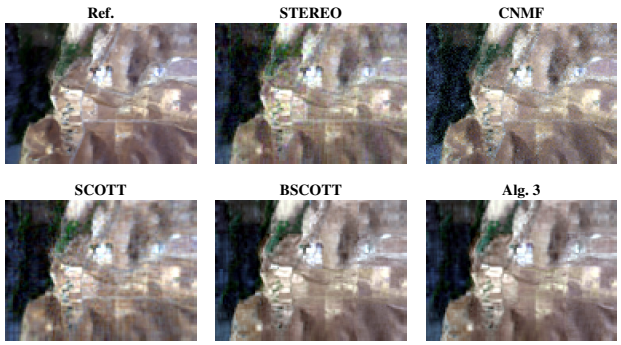

Fig. 4. False color plots of reconstructed SRI, Isabella Lake.

In both examples, the CP-based algorithm STEREO usually yielded the best metrics. It was followed by BSCOTT and Algorithm 2, that gave competitive results both visually and in terms of metrics. The results of SCOTT were slightly lower, and comparable to those of other matrix-based approaches, such as CNMF. Algorithm 2 had the lowest computation time for our implementation, that was comparable to that of BSCOTT.

4.3. Pansharpening

We also addressed the pansharpening problem, which consists in fusion between an HSI $\mathcal{Y}_H$ and a panchromatic image $\mathcal{Y}_P$ of the same scene. The panchromatic image has one single spectral band corresponding to the full spectral range of the SRI, hence $\mathcal{Y}_P \in \mathbb{R}^{I\times J}$ and $\mathbf{P}_3 \in \mathbb{R}^{K\times 1}$.

The CP approach is unfeasible for this problem, since it would require CP-rank $F = 1$. Nevertheless, we could still compare our approach to other matrix-based algorithms. For Algorithm 2 and BSCOTT, we took ranks $(12, 12, 1)$ and $(25, 25, 1)$ for the $[4, 4]$ and $[2, 2]$ splitting scenarios, respectively. The results were shown in Table 4 and Figure 5, respectively.

Although the performance of our algorithm were usually lower than that of matrix methods, it was comparable to that of BSCOTT and even slightly outperformed SCOTT, at a lower computational cost. Visually, the quality of the reconstruction was slightly worse and the block pattern was visible. Despite these difficulties, the proposed approach was still able to recover the main details of the image.

Algorithm	R-SNR	CC	SAM	ERGAS	Time
Best	∞	1	0	0	0
SCOTT	9.8176	0.7473	15.991	10.076	2.377
BSCOTT $[4,4]$	11.3274	0.8182	13.091	8.5843	**0.2171**
Alg. 3 $[4,4]$	11.2986	0.8172	13.221	8.6121	**0.1847**
Alg. 3 $[2,2]$	10.7388	0.7941	14.454	9.1427	0.2598
CB-STAR	10.0096	0.7600	15.847	9.8705	19.873
CNMF	**17.1372**	**0.9529**	**5.833**	**4.6912**	2.610
GLP-HS	**16.1599**	**0.9419**	**6.538**	**5.1209**	10.42
HySure	11.7478	0.8437	13.233	8.1935	121.5
FUSE	16.074	0.9408	6.829	5.1833	0.8814

Table 4. Reconstruction metrics, Pavia University.

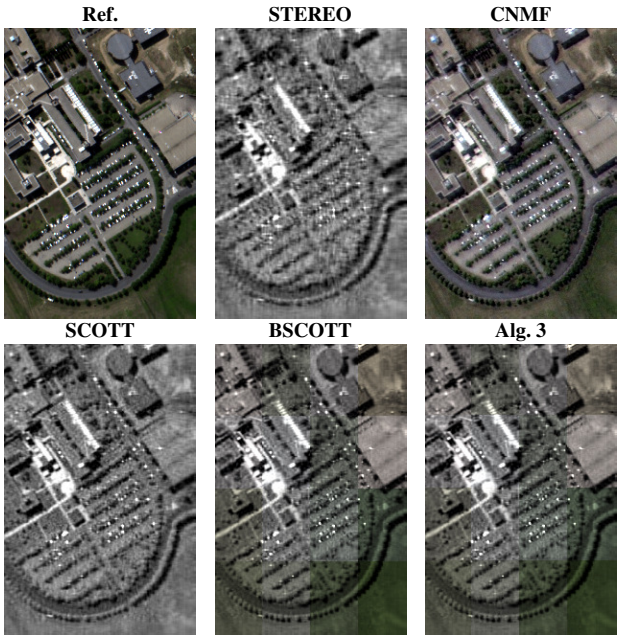

Fig. 5. False color plots of reconstructed SRI, Pavia University.

5. CONCLUSION

In this paper, we proposed an improved Tucked-based algorithm for the HSR problem. It is able to exploit the low-resolution information in the model, to improve the reconstruction performance. We gave exact recovery guarantees for the SRI in the noiseless case by the proposed approach We showed that this new algorithm reaches state-of-the-art performance for HSR and pansharpening at a low complexity, in spite of restrictions on the multilinear ranks. The good reconstruction performance of the proposed method naturally raise the question of its statistical efficiency. Since the algorithm fully exploits the coupling constraints between the Tucker factors, we expect it to be asymptotically efficient for tensor reconstruction. This matter will be investigated in future works.

6. REFERENCES

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

[2] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral

and multispectral data fusion: A comparative review of the recent literature," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 29–56, 2017.

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

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

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

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

[7] 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.

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

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

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

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

[12] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. M. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE J. Sel. Topics Signal Process.*, 2021.

[13] R. Dian, L. Fang, and S. Li, "Hyperspectral image superresolution via non-local sparse tensor factorization," *IEEE Conf. on Comput. Vision and Pattern Recogn. (CVPR)*, pp. 5344– 5353, 2017.

[14] 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.

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

[16] 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.

[17] V. Simoncini, "Computational methods for linear matrix equations," *SIAM Review*, vol. 58, no. 3, pp. 377–441, 2016.

[18] G. H. Golub, A. Hoffman, and G. W. Stewart, "A generalization of the eckart-young-mirsky matrix approximation theorem," *Linear Algebra and its applications*, vol. 88, pp. 317–327, 1987.

[19] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[20] A. J. Laub, *Matrix Analysis For Scientists And Engineers*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

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

[22] 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.

[23] R. E. Roger and J. F. Arnold, "Reliably estimating the noise in AVIRIS hyperspectral images," *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1951–1962, 1996.

[24] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogrammetric Eng. and Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

[25] Q. Wei, J. M. 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.