HYPERSPECTRAL SUPER-RESOLUTION VIA COUPLED TENSOR FACTORIZATION: IDENTIFIABILITY AND ALGORITHMS

Charilaos I. Kanatsoulis, Xiao Fu†, Nicholas D. Sidiropoulos⋆, and Wing-Kin Ma‡*

*Department of ECE, University of Minnesota, Minneapolis, MN, USA
†School of EECS, Oregon State University, Corvallis, OR, USA
⋆Department of ECE, University of Virginia, Charllotesville, VA, USA
‡Department of EE, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

ABSTRACT

This work focuses on the problem of fusing a hyperspectral image (HSI) and a multispectral image (MSI) to produce a super-resolution image that admits high spatial and spectral resolutions. Existing algorithms are mostly based on joint low-rank factorization of the matricized HSI and MSI. This framework is effective to some extent, but several challenges remain. First, it is unclear whether or not the super-resolution image is identifiable in theory under this framework, while identifiability usually plays an essential role in such estimation problems. Moreover, most algorithms assume that the degradation operators from the super-resolution image to the HSI and MSI are known or easily estimated – which is hardly true in practice. In this work, we propose a novel coupled tensor decomposition method that can effectively circumvent these issues. The proposed approach guarantees the identifiability of the super-resolution image under realistic conditions. The method can work even without knowing the spatial degradation operator, which could be hard to accurately estimate in practice. Simulations using AVIRIS Cuprite data are employed to demonstrate the effectiveness of the proposed approach.

Index Terms— Hyperspectral imaging, multispectral imaging, super-resolution, image fusion, tensor decomposition, identifiability, blind deconvolution

1. INTRODUCTION

Fusing images sensed by multiple sensors is of great interest to many applications [1, 2]. In remote sensing, an important fusion problem is to integrate the information from a hyperspectral image (HSI) and a multispectral image (MSI), both of which cover the same object or area. HSIs have very high spectral resolution but relative coarse spatial resolution, while MSIs have fine spatial resolution but low spatial resolution. By fusing an HSI and an MSI, a super-resolution image can be obtained, which can greatly help many analytical tasks such as spectral unmixing and object detection.

Many HSI-MSI fusion methods in the literature treat the problem from a low-rank matrix factorization viewpoint [3, 4, 5, 6, 7, 8, 9]. Intuitively, every pixel of the HSI or MSI can be modeled as a convex combination of spectral signatures of several materials (or *endmembers*). By (jointly) unmixing HSI and MSI to such representation, high-resolution endmembers (from HSI) and the corresponding high-resolution spatial distribution (from MSI) can be estimated. Combining the two, a super-resolution image can be obtained. There are several variations of this basic idea, such as using sparse representations, other factorization models (e.g., SVD), and Bayesian approaches [3, 6, 8]. The matrix factorization approach has proven effective, but there are two major shortcomings. First, it is unclear if the joint HSI and MSI factorization criteria proposed in [3, 4, 5, 6, 7] can guarantee the recovery of the super-resolution image; i.e., there is no assured identifiability. Second, the approaches usually assume that the degradation operators from the super-resolution image to the HSI and MSI are known [4, 5, 6, 7] or can be estimated from data [3]. Assuming precise knowledge of the degradation operators, especially the spatial degradation operator (which involves kernel function and hyper-parameter selection), is often impractical. Without calibration/training data, it is unclear if the kernel is blindly identifiable.

In this work, we offer an alternative solution to the HSI-MSI fusion problem. Our method starts from the fact that both HSI and MSI images are space-space-spectrum "cubes", and thus are naturally three-way tensors [10]. One nice property of tensors is that *any* tensor admits a canonical polyadic decomposition (CPD), and the decomposition is essentially unique under quite mild conditions. Hence, we propose a coupled tensor factorization approach for the HSI-MSI fusion problem. Leveraging the uniqueness of three-way tensor factorization, we show that this method can provably identify the super-resolution image. We further show that even when the spatial degradation operator, which is in general hard to estimate, is unknown, the proposed approach can still guarantee the identification of the super-resolution image, with slight modifications. Numerical simulations using HSI and MSI images that are degraded from Cuprite data show that the proposed approach is very promising for the HSI-MSI fusion task.

2. PROBLEM STATEMENT AND BACKGROUND

Let us consider a hyperspectral image $\underline{\boldsymbol{Y}}_H \in \mathbb{R}^{I_H \times J_H \times K_H}$, where I_H and J_H denote dimensions that span the spatial domain and K_H denotes the number of spectral bands. Similarly, we denote a multispectral image cube as $\underline{\boldsymbol{Y}}_M \in \mathbb{R}^{I_M \times J_M \times K_M}$, where I_M, J_M and K_M denote the dimensions of two spatial and one spectral coordinates, respectively. We assume that the two images are aligned so that they describe the same region in the spatial domain. HSIs typically have hundreds of spectral bands while MSIs have less than 20; i.e., $K_M \ll K_H$ in general. On the other hand, MSIs have much finer resolution in the spatial domain relative to HSIs – i.e., $I_H J_H \ll I_M J_M$ typically holds.

Our goal is to integrate HSI and MSI so that a super-resultion image cube is obtained. That is, we aim at obtaining a $\underline{\boldsymbol{Y}}_S \in \mathbb{R}^{I_M \times J_M \times K_H}$ that has the spatial resolution of the MSI and the spectral resolution of the HSI. As mentioned, this task is of great interest to image processing and analytics in geoscience, food in-

Supported in part by U.S. NSF grants ECCS-1608961, IIS-1447788, and IIS-1704074.

spection, anomaly detection and target recognition.

Matrix Factorization-Based Approach. Many recent methods for HSI-MSI fusion make use of the fact that the matricized HSI and MSI data are low-rank matrices and come up with low-rank decomposition based methods to handle the fusion problem [3, 4, 5, 6, 7, 8, 9]. Specifically, consider

$$\boldsymbol{Y}_H = [\underline{\boldsymbol{Y}}_H(1,1,:), \ldots, \underline{\boldsymbol{Y}}_H(I_H, J_H, :)] \in \mathbb{R}^{K_H \times I_H J_H}, \quad (1)$$

where $\underline{\boldsymbol{Y}}_H(i,j,:) \in \mathbb{R}^{K_H}$ is a vector that is formed by taking the (i,j)th pixel of the HSI. By the linear mixture model (LMM) that is commonly adopted in hyperspectral imaging, $\boldsymbol{Y}_H \approx \boldsymbol{G}_H \boldsymbol{S}_H^T$, where $\boldsymbol{G}_H \in \mathbb{R}^{K_H \times R}$, $\boldsymbol{S}_H^T \in \mathbb{R}^{R \times I_H J_H}$, $R \ll \min\{I_H J_H, K_H\}$, and $\mathbf{1}^T \boldsymbol{S}_H^T = \mathbf{1}^T$ and $\boldsymbol{S}_H \geq \mathbf{0}$ holds. This low-rank factorization model is based on physical modeling of the pixels of HSIs: A pixel (point spectrum) $\boldsymbol{Y}_H(:,\ell)$ is modeled as a weighted sum of the spectral signatures of several materials (or endmembers) that are present in that pixel. This is a widely accepted model. By similar arguments, we have $\boldsymbol{Y}_M \approx \boldsymbol{G}_M \boldsymbol{S}_M^T$ with rank R, where the $\boldsymbol{Y}_M$ is the matricized MSI. Hence, the matricized super-resolution image can be sythesized by $\boldsymbol{Y}_S \approx \boldsymbol{G}_H \boldsymbol{S}_M^T$. Here, again, $\boldsymbol{Y}_S$ is obtained by applying the operation in (1) to $\underline{\boldsymbol{Y}}_S$.

The common assumption that is adopted in matrix factorization approaches is that there exist two linear operators $\boldsymbol{P}_H \in \mathbb{R}^{I_M J_M \times I_H J_H}$ and $\boldsymbol{P}_M \in \mathbb{R}^{K_M \times K_H}$ such that $\boldsymbol{Y}_H = \boldsymbol{Y}_S \boldsymbol{P}_H^T$ and $\boldsymbol{Y}_M = \boldsymbol{P}_M \boldsymbol{Y}_S$ can be obtained from the super-resolution image $\boldsymbol{Y}_S$ via linear transformations. Consequently, we have $\boldsymbol{Y}_H = \boldsymbol{G}_H (\boldsymbol{P}_H \boldsymbol{S}_M)^T$ and $\boldsymbol{Y}_M = (\boldsymbol{P}_M \boldsymbol{G}_H) \boldsymbol{S}_M^T$. Then, $\boldsymbol{G}_H$ and $\boldsymbol{S}_M$ can be estimated via jointly factoring $\boldsymbol{Y}_H$ and $\boldsymbol{Y}_M$ following the desribed model. The above is the common basic idea behind the approaches in [3, 4, 5, 6, 7, 8, 9], where various low-rank models and factorization criteria are employed to enhance performance.

Challenges. The matrix factorization-based approaches for HSI-MSI fusion are fairly effective and considered state of the art. Nevertheless, two key theoretical and practical challenges remain.

First, existing methods rarely consider the identifiability of $\boldsymbol{Y}_S$. Note that recovering $\boldsymbol{G}_H \boldsymbol{S}_M$ from compressed measurements $\boldsymbol{Y}_H$ and $\boldsymbol{Y}_M$ can be quite ill-posed, since there are many solutions that satisfy $\boldsymbol{Y}_H = \boldsymbol{G}_H \boldsymbol{S}_H^T$ and $\boldsymbol{Y}_M = \boldsymbol{G}_M \boldsymbol{S}_M^T$. One could argue identifiability from a matrix sensing viewpoint [11], but identifiability for matrix sensing is guaranteed when the degradation operators are random. In our context, these operators are highly structured – thus known theory cannot answer the question. The coupled factorization approaches with a variety of regularizations [7, 4, 6, 3] may help in practice – but currently lack theoretical guarantees. Note that identifiability is of great interest not only from a theoretical viewpoint, but also often serves as guidance for practitioners to select and design the 'correct' solvers and algorithms – which have been proven very useful and powerful in pertinent problems, such as spectral unmixing [12].

Second, most matrix factorization-based fusion algorithms assume that $\boldsymbol{P}_H$ and $\boldsymbol{P}_M$ are accurately known or can be easily estimated, which is hardly true in practice. The matrix $\boldsymbol{P}_M$ is relatively easier to obtain. However, modeling $\boldsymbol{P}_H$ is much more difficult. The commonly used model is to represent $\boldsymbol{P}_H$ as an operator that blurs q-by-q overlapping grids in the 2D spatial domain of the super-resolution image and then downsamples one pixel from the blurred grids to form a low spatial resolution image. This process involves several factors that are unknown in practice – e.g., the blurring function and the grid size. There are approaches in the literature, e.g., [3], that propose to estimate the degradation operators from data. But these approaches involve a series of structural assumptions (e.g., smoothness and sparsity) on the operators and hyperparameter tuning.

3. PROPOSED APPROACH

In this section, to circumvent the above challenges, we propose a tensor-based approach to handle the MSI-HSI fusion problem.

3.1. Tensor Algebra Preliminaries

Our method heavily uses tensor algebra. To facilitate our discussion, we briefly review some key concepts that will be used in our approach. A three-way tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$ can be considered as a three-way array whose elements are indexed by i, j, k. A tensor can always be 'explained' by the so-called canonical polyadic decomposition (CPD) model, i.e., $\underline{\boldsymbol{X}}(i,j,k) = \sum_{f=1}^{F} \boldsymbol{A}(i,f)\boldsymbol{B}(j,f)\boldsymbol{C}(k,f)$, with a proper F which we refer to as the tensor rank or CPD rank [10], 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 called the low-rank factors of the three-way tensor. Since a three-way tensor can be fully characterized by its low-rank factors, we sometimes use the notation $\underline{\boldsymbol{X}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ to represent the tensor.

One nice property of tensors is that the CPD model is essentially unique even when F is much larger than $\max\{I, J, K\}$. For example, we have the following theorem:

Theorem 1 *[13] Let $\underline{\boldsymbol{X}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ with $\boldsymbol{A} : I \times F$, $\boldsymbol{B} : J \times F$, and $\boldsymbol{C} : K \times F$. Assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from some continuous distributions. Also assume $I \geq J \geq K$ without loss of generality. If $F \leq 2^{\lfloor \log_2 J \rfloor + \lfloor \log_2 K \rfloor - 2}$, then the decomposition of $\underline{\boldsymbol{X}}$ in terms of $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ is essentially unique, almost surely. In the case $K \geq 2^{\lfloor \log_2 J \rfloor + \lfloor \log_2 K \rfloor - 2}$ and $J + K > 3$ the above bound can be relaxed to $F \leq \min(I, (J-1)(K-1))$.*

Here, essential uniqueness means that if $\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{C}}$ also satisfy $\underline{\boldsymbol{X}} = [\![\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{C}}]\!]$, we can only have $\boldsymbol{A} = \tilde{\boldsymbol{A}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1$, $\boldsymbol{B} = \tilde{\boldsymbol{B}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_2$, and $\boldsymbol{C} = \tilde{\boldsymbol{C}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3$, where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Lambda}_i$ is a full rank diagonal matrix such that $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \boldsymbol{I}$. One can see that the uniqueness condition is rather mild: For example, if one has a $80 \times 80 \times 80$ tensor, then it admits an essentially unique CPD representation if $F \leq 1024$. Note that this is far more relaxed compared to uniqueness conditions for matrix factorization, where nonnegativity, sparsity, and geometric conditions are needed and the rank has to be lower than the outer dimensions of the matrix [14, 15].

Two useful operations for tensor are matricization and taaking a mode product. The operation in (1) is in fact one way to matricize a three-way tensor. In this work, the transpose of operation in (1) is used and the matricized $\underline{\boldsymbol{X}}$ has the following form: $\boldsymbol{X} = (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^T \in \mathbb{R}^{I_H J_H \times K_H}$, where $\odot$ denotes the Kronecker product. The mode product operator is essentially mutiplying a matrix to all the slabs of a tensor in one mode (note that a three-way tensor has three different types of slabs, i.e., front, horizontal, and vertical slabs, and thus three modes). Therefore $\underline{\tilde{\boldsymbol{X}}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 \times_3 \boldsymbol{P}_3$ leads to a matricized form $\tilde{\boldsymbol{X}} = (\boldsymbol{P}_2\boldsymbol{B} \odot \boldsymbol{P}_1\boldsymbol{A})(\boldsymbol{P}_3\boldsymbol{C})^T$; see details in [10, 16].

3.2. Degradation as Mode Products

We start with the case where $\boldsymbol{P}_H$ and $\boldsymbol{P}_M$ are both known and show how to formulate the fusion problem as a coupled tensor factorization problem. Then, we will discuss identifiability issues including the case where $\boldsymbol{P}_H$ is unknown.

Note that the HSI, MSI, and the super-resolution image are all naturally modeled as three-way space-frequency tensors. The super-resolution tensor $\underline{\boldsymbol{Y}}_S$ admits a CPD model, i.e., $\underline{\boldsymbol{Y}}_S = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ for appropriate F. Let us assume that the spatial degradation to

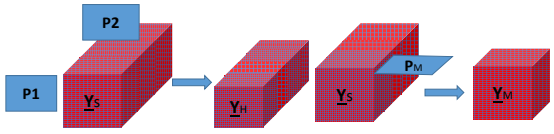

Fig. 1: Illustration of degradation from the super-resolution image to the HSI and MSI, respectively.

the HSI can be modeled by $\underline{\boldsymbol{Y}}_H(:,:,k) = \boldsymbol{P}_1 \underline{\boldsymbol{Y}}_S(:,:,k)\boldsymbol{P}_2^T$, $k = 1,\ldots,K_H$, where $\boldsymbol{P}_1$ is a blurring and dimensionality-reducing matrix that operates on the row dimension of each slab, and $\boldsymbol{P}_2$ does the same on the column dimension. Such separability is commonly assumed in the matrix factorization approaches, and it makes sense in practice. For example, the commonly used 2D Gaussian blurring and downsampling procedure, modeled by $\boldsymbol{P}_H\boldsymbol{Y}_S$ in the matricized form [4, 5, 6, 7], is a separable operation and can be 'decomposed' to an one-dimensional Gaussian convolution plus downsampling across rows and another one-dimensional Gaussian convolution plus downsampling across columns of $\boldsymbol{Y}_S(:,:,k)$'s. This is equivalent to a 2-D blurring and downsampling procedure using two operators $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ that are applied to the rows and columns, i.e., $\underline{\boldsymbol{Y}}_H = \underline{\boldsymbol{Y}}_S \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2$ and $\boldsymbol{P}_H = \boldsymbol{P}_2 \otimes \boldsymbol{P}_1$, where $\otimes$ denotes the Kronecker product. In the matricized form, we have $\boldsymbol{Y}_H = ((\boldsymbol{P}_2\boldsymbol{B}) \odot (\boldsymbol{P}_1\boldsymbol{A}))\boldsymbol{C}^T$. Similarly, it is also readily seen that $\boldsymbol{Y}_M = (\boldsymbol{B} \odot \boldsymbol{A})(\boldsymbol{P}_M\boldsymbol{C})^T$.

3.3. Coupled Tensor Factorization for Super-resolution

Let us first consider the case where $\boldsymbol{P}_M$ and $\boldsymbol{P}_H$ are known. In this case, we propose to employ the following formulation for HSI-MSI super-resolution:

$$\underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}}{\text{minimize}} \ \|\underline{\boldsymbol{Y}}_H - [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]\|_F^2 + \lambda\|\underline{\boldsymbol{Y}}_M - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]\|_F^2, \tag{2}$$

where $\lambda > 0$ is a pre-selected parameter that balances the two terms. Specifically, we jointly decompose the HSI tensor and the MSI tensor to estimate $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ using the above. Then, we can reconstruct the super-resolution tensor using the latent factors, i.e., $\hat{\underline{\boldsymbol{Y}}}_S(i,j,k) = \sum_{f=1}^F \hat{\boldsymbol{A}}(i,f)\hat{\boldsymbol{B}}(j,f)\hat{\boldsymbol{C}}(k,f)$. The algorithm for solving Problem (2) is presented in Algorithm 1. We refer to this algorithm as *tensor-based super-resolution* (`SuperTensor` for short). In the algorithm, $\boldsymbol{H}_i$ and $\boldsymbol{M}_i$ denote the ith mode matricization of the HSI and MSI, respectivley. The idea is to optimize Problem (2) with respect to (w.r.t.) $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ one at a time while fixing the other two, and repeating the process cyclically – which is reminiscent of the classic alternating least squares (ALS) algorithm in the tensor literature. Note that each subproblem in `SuperTensor` is an unconstrained least squares problem and thus can be solved rather efficiently. Also note that various other optimization approaches can be employed, but we have to postpone this discussion for the journal version.

We also consider the case where the spatial degradation operators $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are completely unknown. In that case, we propose to employ the following estimator:

$$\underset{\boldsymbol{A},\boldsymbol{B},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{B}},\boldsymbol{C}}{\text{minimize}} \ \left\|\underline{\boldsymbol{Y}}_H - \left[\!\!\left[\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \boldsymbol{C}\right]\!\!\right]\right\|_F^2 + \lambda\|\underline{\boldsymbol{Y}}_M - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]\|_F^2. \tag{3}$$

In the above formulation, we replaced $\boldsymbol{P}_1\boldsymbol{A}$ and $\boldsymbol{P}_2\boldsymbol{B}$ with $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}$, respectively. A similar alternating optimization approach can be employed to handle the criterion. The motivation of the formulation

Algorithm 1: `SuperTensor`

Initialization: λ, F, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$
repeat
  $\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \|\boldsymbol{H}_1 - (\boldsymbol{C} \odot \boldsymbol{P}_2\boldsymbol{B})\boldsymbol{A}^T\boldsymbol{P}_1^T\|_F^2 + \lambda\|\boldsymbol{M}_1 - (\boldsymbol{P}_3\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T\|_F^2$;
  $\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \|\boldsymbol{H}_2 - (\boldsymbol{C} \odot \boldsymbol{P}_1\boldsymbol{A})\boldsymbol{B}^T\boldsymbol{P}_2^T\|_F^2 + \lambda\|\boldsymbol{M}_2 - (\boldsymbol{P}_3\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^T\|_F^2$;
  $\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \|\boldsymbol{H}_3 - (\mathbf{P}_2\boldsymbol{B} \odot \mathbf{P}_1\boldsymbol{A})\boldsymbol{C}^T\|_F^2 + \lambda\|\boldsymbol{M}_3 - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^T\boldsymbol{P}_3^T\|_F^2$;
until Some stopping criterion is met
Reconstruct $\underline{\boldsymbol{Y}}_S$ using $\hat{\underline{\boldsymbol{Y}}}_S(i,j,k) = \sum_{f=1}^F \boldsymbol{A}(i,f)\boldsymbol{B}(j,f)\boldsymbol{C}(k,f)$.

is as we stated in the last subsection: The blurring and downsampling kernel for the spatial degradation is hard to know. Note that if one could identify $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ from (3), one can still reconstruct $\underline{\boldsymbol{Y}}_S$. We should mention that we assume knowledge of $\boldsymbol{P}_M$ for a couple of reasons: First, it is relatively easy to get a reasonable estimate of $\boldsymbol{P}_M$ by simply inspecting the employed wavelengths of the HSI and MSI cameras, respectively – which means there is no hyper-parameter such as the size of the blurring grid as in $\boldsymbol{P}_H$ that is hard to determine in practice. Second, the common factor $\boldsymbol{C}$ in both terms of (3) is essential for coupling the two factorizations together and fixing the permutation and scaling ambiguities that are inherent in tensor decomposition, which we need for reconstruction.

3.4. Identifiability Issues

Regarding the identifiability of the super-resolution image cube, we have the following theorems:

Theorem 2 *Let $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$. Assume without loss of generality that $I_M \geq J_M \geq K_M$. Also assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from some continuous distribution and $(\boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{C}^\star)$ is an optimal solution to Problem (2). Then, $\hat{\underline{\boldsymbol{Y}}}_S(i,j,k) = \sum_{f=1}^F \boldsymbol{A}^\star(i,f)\boldsymbol{B}^\star(j,f)\boldsymbol{C}^\star(k,f)$ recovers the ground-truth $\underline{\boldsymbol{Y}}_S$ almost surely if $F \leq 2^{\lfloor\gamma\rfloor-2}$, where $\gamma = \log_2(J_M K_M)$. In the case $I_M \geq 2^{\lfloor\gamma\rfloor-2}$ and $J_M + K_M > 3$ the above bound can be relaxed to $F \leq \min(I_M, (J_M-1)(K_M-1))$.*

For the case where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are unknown, we also have

Theorem 3 *Let $\underline{\boldsymbol{Y}}_H = [\![\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$. Assume without loss of generality that $I_H \geq J_H \geq K_H$ and $I_M \geq J_M \geq K_M$. Also assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from some continuous distribution and $(\tilde{\boldsymbol{A}}^\star, \tilde{\boldsymbol{B}}^\star, \boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{C}^\star)$ is an optimal solution to Problem (3). Then, $\hat{\underline{\boldsymbol{Y}}}_S(i,j,k) = \sum_{f=1}^F \boldsymbol{A}^\star(i,f)\boldsymbol{B}^\star(j,f)\boldsymbol{C}^\star(k,f)$ recovers the ground-truth $\underline{\boldsymbol{Y}}_S$ almost surely if $F \leq \min\{2^{\lfloor\gamma_1\rfloor-2}, 2^{\lfloor\gamma_2\rfloor-2}\}$, where $\gamma_1 = \log_2(J_M K_M)$ and $\gamma_2 = \log_2(J_H K_H)$. In the case $I_M \geq 2^{\lfloor\gamma_1\rfloor-2}$ and $J_M + K_M > 3$, $2^{\lfloor\gamma_1\rfloor-2}$ can be replaced by $\min(I_M, (J_M-1)(K_M-1))$. Equivalently when $I_H \geq 2^{\lfloor\gamma_2\rfloor-2}$ and $J_H + K_H > 3$, $2^{\lfloor\gamma_2\rfloor-2}$ can be replaced by $\min(I_H, (J_H-1)(K_H-1))$.*

The proofs of the theorems are relegated to a journal version due to space limitations; the idea is to make use of Theorem 1 to characterize the optimal solutions of Problems (2) and (3). To have some concrete sense of the theorems, let consider an example where we are interested in reconstructing a super-resolution image of size $600 \times 512 \times 130$ from an HSI of size $150 \times 128 \times 130$ and an MSI of size $600 \times 512 \times 8$. Then reconstruction is guaranteed under

the assumptions of Theorem 2 or 3 if the CPD rank of the super-resolution image tensor (as well as those of the HSI and MSI) satisfies $F \leq 1024$. This is in general easy to satisfy (approximately) in practice. In table 1, we show the normalized mean squared error (NMSE) of using a CPD model to reconstruct a subimage of the AVIRIS Cuprite image [17]. One can see that for all the tested ranks, the reconstruction error is rather small. Note that under the tested ranks, the CPD model is unique. This means that using a unique CPD model to approximate HSI and MSI cubes is very reasonable.

Table 1: The NMSE of using a CPD model to approximate a subimage of the AVIRIS Cuprite data that is of size $512 \times 614 \times 187$.

rank	300	400	500	600	700	800
NMSE	0.019	0.016	0.0142	0.0131	0.0123	0.0116

4. SIMULATIONS

In this section, we use a real available HSI as a reference (i.e., to act as the super-resolution image) and synthesize a HSI and MSI following the so-called Wald's protocol [18] that is widely adopted in the HSI-MSI fusion literature. Specifically, we model the degradation process from super-resolution to the HSI as a combination of image blurring by a 9×9 Gaussian kernel and down-sampling the result by a factor of $d = 16$ as described previously. To obtain the MSI, the spectral response $\boldsymbol{P}_M$ is modeled as a matrix that picks and averages certain bands of the super-resolution spectrum. The dataset used in the experiments is the Cuprite HSI downloaded by the AVIRIS platform. It represents geological features in 187 bands of spatial resolution 512×614, i.e. $\underline{\boldsymbol{Y}}_S \in \mathbb{R}^{512\times614\times187}$. Then, $\underline{\boldsymbol{Y}}_H \in \mathbb{R}^{128\times152\times187}$ and $\underline{\boldsymbol{Y}}_M \in \mathbb{R}^{512\times614\times6}$ are the produced.

The baseline algorithms used for comparison are: `Naive`, where each pixel of the HSI is replicated as many times as needed to fit the super spatial resolution, `Bayes-Naive` [5] and `FUMI` [7], among which `FUMI` is considered state-of-the-art. The algorithm in [19] was also tested but failed to operate due to memory overflow. To evaluate performance, we adopt popularly used metrics as defined in [20, 1], namely cross correlation (CC), spectral angle mapper (SAM), and relative dimensional global error (ERGAS), and the reconstruction normalized mean squared error (NMSE). In a nutshell, high values of CC and SAM and low values of ERGAS and NMSE correspond to good fusion performance.

All simulations were performed in Matlab on a Linux server with 3.6GHz cores and 32GB RAM. We implemented two versions of `SuperTensor`: Besides the original one, we also implement `SuperTensor` with a commonly employed pre-processing in tensor decompositions. Specifically, we apply mode SVD to the third mode of the HSI tensor and reduce dimension of that mode to R and then feed the reduced-dimension tensor to `SuperTensor`. Note that factors of the original tensor can be reconstructed from the factors of the dimension-reduced tensor if the third-mode matricization has low matrix rank. This way, the computational complexity of tensor factorization can be reduced and empirically better results are often observed; see [10]. We call this implementation `SuperTensor` with pre-processing (p.p.). We initialize `SuperTensor` by the CPD of $\underline{\boldsymbol{Y}}_M$, which is computed using `Tensorlab` [21], run for maximum 20 iterations. In all the simulations, we fix the λ parameter to be 10^{-2}. The stopping criterion of `FUMI` is set to be the relative cost error being less than 10^{-4} as suggested in [7].

Table 2 shows the performance of the algorithms . The rank used for the full tensor decompositions is $F = 750$ and the rank of the low rank matrix model is $R = 10$. It is important here to note that $F = 750$ was chosen following Theorems 2-3 and also Table 1. One can see that both `SuperTensor`, `SuperTensor` w/ p.p. and `FUMI` produce comparably good super-resolution images, and that `SuperTensor` w/ p.p. works best from an accuracy point of view, while the runtime of the `SuperTensor` algorithms is much less than that of `FUMI`. Specifically, both `SuperTensor` versions use 1/3 the runtime of `FUMI`. Also note that `SuperTensor` offers identifiability guarantees, while `FUMI` in general does not.

Table 2: Performance of the algorithms assuming the degradation operators are known.

Algorithm	**NMSE**	**CC**	**SAM**	**ERGAS**	**runtime** (min)
`SuperTensor`	0.0164	0.99223	0.88647	0.46708	10
`SuperTensor` (w/ p.p.)	**0.0149**	**0.99349**	0.81903	**0.4359**	10
`Bayes-Naive`	0.0309	0.97336	1.0663	0.8807	xx
`FUMI`	0.0149	0.9933	**0.80617**	0.4441	30
`Naive`	0.0646	0.88234	1.228	1.7136	xx

The second set of experiments examines the case where the degradation model from super-resolution to HSI is not accurately known. In particular we consider a scenario where $\underline{\boldsymbol{Y}}_H$ is produced by $\underline{\boldsymbol{Y}}_s$ after 9×9 Gaussian blurring and downsampling, but the baseline algorithms falsely assume 11×11 or 13×13 averaging-based blurring. We implement `SuperTensor` with the formulation in (3), which we refer to as `SuperTensor-B` where 'B' stands for 'Blind'. Table 3 shows the performance of the algorithms. One can see that the proposed method outperforms the baselines on three out of four accuracy metrics, which shows that the method is really robust to model mismatch. In terms of runtime, the result is similar with the previous case – the proposed method is 3 times faster compared to `FUMI`.

Table 3: Performance of the algorithms when the spatial degradation model is not accurately known.

Algorithm	**NMSE**	**CC**	**SAM**	**ERGAS**	**runtime (min)**
`SuperTensor-B`	**0.0219**	**0.98776**	1.1367	0.**59665**	x
`Bayes-Naive` (11×11)	0.0324	0.97006	1.0665	0.91537	x
`Bayes-Naive` (13×13)	0.0330	0.96897	1.0666	0.92849	x
`FUMI` (11×11)	0.0307	0.9757	1.1236	**0.82304**	x
`FUMI` (13×13)	0.0329	0.97228	1.1558	0.87747	x
`Naive`	0.0646	0.88234	1.228	1.7136	x

5. CONCLUSION

In this work, we proposed a new coupled tensor factorization based framework for the HSI-MSI fusion problem. Compared to the popular approaches that are mostly based on coupled matrix factorization, the proposed framework enjoys several favorable features: The method can provably identify the super-resolution tensor under realistic conditions, and is also friendly for algorithm design. In addition, this framework can easily accomandate scenarios where the spatial degradation model is unclear or inaccurately estimated, which is often the case in practice. Simulations using the AVIRIS Curpite image data show that the proposed method is effective and efficient in fusing HSI and MSI under different scenarios.

6. REFERENCES

[1] L. Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.

[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 *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[3] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[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, 2012.

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

[6] 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.

[7] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7236–7249, 2016.

[8] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.

[9] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1409–1413.

[10] 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.

[11] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 665–674.

[12] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, Jan 2014.

[13] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 3, pp. 1018–1037, 2012.

[14] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6254–6268, 2016.

[15] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of nonnegative matrix factorization," *arXiv preprint arXiv:1709.00614*, 2017.

[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[17] G. Swayze, R. N. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada," 1992.

[18] 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, pp. 691–699, 1997.

[19] 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.

[20] 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.

[21] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab v3. 0, march 2016," *URL: http://www. tensorlab. net*.