

Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization

Shutao Li, *Senior Member, IEEE*, Renwei Dian, *Student Member, IEEE*, Leyuan Fang, *Senior Member, IEEE*, and José M. Bioucas-Dias, *Fellow, IEEE*

Abstract—Fusing a low spatial resolution hyperspectral image (LR-HSI) with a high spatial resolution multispectral image (HR-MSI) to obtain a high spatial resolution hyperspectral image (HR-HSI) has attracted increasing interest in recent years. In this paper, we propose a coupled sparse tensor factorization (CSTF) based approach for fusing such images. In the proposed CSTF method, we consider an HR-HSI as a three-dimensional tensor and redefine the fusion problem as the estimation of a core tensor and dictionaries of the three modes. The high spatial-spectral correlations in the HR-HSI are modeled by incorporating a regularizer which promotes sparse core tensors. The estimation of the dictionaries and the core tensor are formulated as a coupled tensor factorization of the LR-HSI and of the HR-MSI. Experiments on two remotely sensed HSIs demonstrate the superiority of the proposed CSTF algorithm over current state-of-the-art HSI-MSI fusion approaches.

Index Terms—super-resolution, fusion, hyperspectral imaging, coupled sparse tensor factorization.

I. INTRODUCTION

Hyperspectral imaging is an emerging modality where a camera acquires images from a scene across a number of different wavelengths. The very high spectral resolution and coverage of hyperspectral images (HSIs) enable a precise identification of the materials present in the scene, which underlies a large number of remote sensing [1]–[4] and computer vision applications [5], [6]. However, since the photons emitted by the sun are spread over many spectral bands, the spatial resolution has to be decreased in order to maintain the number of photons in each band above a minimum value. In this way, the signal-to-noise-ratio (SNR) due to the Poisson noise is kept above a minimum acceptable value [7]. As a result, the spatial resolution of HSIs is often poor. Compared with hyperspectral imaging sensors, the existing multispectral imaging sensors capture MSIs with much higher spatial resolution and SNR [8]. Therefore, the low spatial resolution hyperspectral images (LR-HSIs) are often fused with high spatial resolution multispectral images (HR-MSIs) to reconstruct hyperspectral

This paper was supported by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, and by the Portuguese Science and Technology Foundation under Projects UID/EEA/50008/2013 and ERANETMED/0001/2014. (*Corresponding author:* Renwei Dian.)

S. Li, R. Dian, and L. Fang are with the College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China (e-mail: shutao_li@hnu.edu.cn; drw@hnu.edu.cn; fangleyuan@gmail.com).

José M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

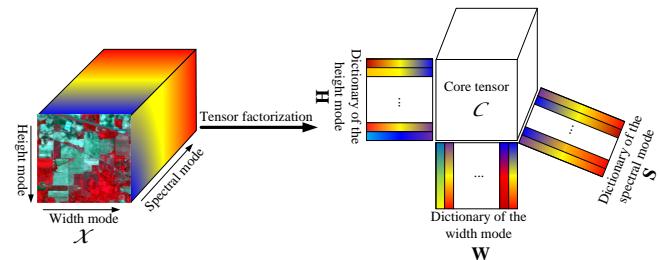


Fig. 1. Illustration of tensor decomposition of the HR-HSI.

images with high spatial resolution (HR-HSIs). This procedure is referred to as HSI super-resolution or HSI-MSI fusion.

Pan-sharpening [9] is a class of seminal fusion methods which infer a high spatial resolution multispectral image from a low spatial resolution version and a panchromatic (PAN) image (gray image). Representative pan-sharpening approaches include the sparse representation based methods [10], [11], the variational methods [12], [13], and the morphological methods [14], etc. These methods can often be extended to the fusion of LR-HSIs and HR-MSIs. For example, Chen *et al.* [15] divide the spectrum of the LR-HSI into several regions and then fuse the LR-HSI with HR-MSI in each region using a pan-sharpening algorithm. For an extensive comparison of hyperspectral pan-sharpening, see [16].

In spatial-spectral image fusion, a class of methods exploits the Bayesian framework [17]–[19] to fuse a LR-HSI with a HR-MSI. Based on prior knowledge and on the observation model, these Bayesian fusion methods build the posterior distribution, which is the Bayesian inference engine. In [17], a maximum a posteriori (MAP) based method is proposed for the fusion of an LR-HSI and an HR-MSI. The work [18] fully exploits the observations model and introduces a Sylvester-based solver, which computes the fused image very efficiently. In [19], a Bayesian sparse representation based spectral linear mixing model is introduced to solve the fusion problem.

Recently, matrix factorization has been actively investigated to fuse pairs of LR-HSIs and HR-MSIs [20]–[28]. Assuming that the HR-HSI only contains a small number of pure spectral signatures [29], these approaches first unfold the HR-HSI as a matrix and then factor it into a basis matrix and a coefficient matrix. The work [20] learns the spectral basis from the LR-HSI with a sparse prior and then conducts sparse coding on the HR-MSI to estimate the coefficient matrix. Instead of estimating spectral basis in advance and keeping them fixed, works [21]–[23] alternately update the spectral basis and the

coefficients, with non-negative constraints, which yields more accurate reconstructions. Aiming at better inferences works [24]–[28] also exploit the spatial similarities typical of HR-HSIs to solve the fusion problem. For example, the work [24] adopts a simultaneous greedy pursuit algorithm to compute the coefficient matrix for each local patch. In this way, the local HR-HSI similarities (i.e., nearby pixels are likely to represent the same materials) are exploited. The work [27] proposes a nonnegative dictionary-learning algorithm to learn the spectral basis and uses a structured sparse coding approach to estimate the coefficient matrix. In this way, the HR-HSI non-local spatial similarities are exploited yielding state-of-the-art performance. In general, all the above matrix factorization based methods need to unfold the three-dimensional data structures into matrices. Although the unfolding operation does not change information of three-dimensional data, the fully exploitation of the spatial-spectral correlations of the HSIs is not easily accomplished by the matrix factorization based methods.

More recently, the work [30] proposes a tensor factorization based method. Different from the matrix factorization approaches, the unknown HR-HSI is approximated by a tensor with four factors: a core tensor multiplied (n-mode products) by dictionaries (factor matrices) of the three modes. The proposed method starts by clustering similar HSI cubes together; then, the dictionaries are learned for each mode and the core tensor for each HR-HSI cube is estimated by conducting sparse coding on the learned dictionaries. Although this method produces good results, the LR-HSI is only used for learning the spectral dictionary, and it is not used in the estimation of core tensor.

Since HSIs can be naturally represented as a three-dimensional tensor, it may be better to deal with them from the viewpoint of tensors. Recently, tensor-based methods have been successfully applied to multi-frame data restoration [31]–[34], recognition [35], [36], and unmixing [37], etc.

As one of the most effective tensor decomposition methods, Tucker decomposition [38] factors a tensor as a core tensor multiplied by the factor matrix along each mode. Inspired by the above works, a novel coupled sparse tensor factorization (CSTF) based approach is herein proposed for the fusion of an LR-HSI and an HR-MSI. Based on the Tucker decomposition, we consider the HR-HSI as the three-dimensional tensor, which can be approximated by a core tensor multiplied by dictionaries of three modes, as shown in Fig. 1. In this way, the problem of HSI-MSI fusion is formulated as the estimation of the core tensor and dictionaries of three modes from a LR-HSI and an HR-MSI. In such a problem, the dictionaries of the width and of the height modes represent the spatial information of the HR-HSI, and the dictionary of the spectral mode represents the spectral information. The core tensor models the relations among the dictionaries of three modes.

HSIs are low-rank and self-similar [26], [28]. These characteristics are strong forms of prior knowledge that we model by imposing a sparse prior on the core tensor. The spectral and spatial information of the HR-HSI exists mainly in the LR-HSI and the HR-MSI, respectively. Therefore, we formulate the estimation of three dictionaries and core tensor as a coupled

sparse tensor factorization of the LR-HSI and HR-MSI. The factorization steps are iterated until convergence. In this way, dictionaries and core tensor are all updated in each iteration yielding accurate estimates.

The remainder of this paper is organized as follows. In Section II, we give a basic overview on tensors. Section III formulates the HSI-MSI fusion. The proposed CSTF approach for HSI-MSI fusion is introduced in Section IV. In Section V, experimental results on two remote sensing HSIs are presented. Conclusions and future research directions are given in Section VI.

II. PRELIMINARIES ON TENSORS

A N-dimensional tensor is denoted as $\mathcal{M} \in \mathbb{R}^{I_1 \times I_2, \dots, \times I_N}$. The elements of \mathcal{M} are denoted as $m_{i_1 i_2, \dots, i_N}$, where $1 \leq i_n \leq I_n$. The n-mode unfolding vectors of tensor \mathcal{M} are the I_n -dimensional vectors obtained from \mathcal{M} by changing index i_n , while keeping the other indices fixed. The n-mode unfolding matrix $\mathbf{M}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2, \dots, I_{n-1} I_{n+1}, \dots, I_N}$ is defined by arranging all the n-mode vectors as the columns of the matrix [39].

The product of two matrices can be generalized to the multiplication of a tensor and a matrix. The n-mode product of the tensor $\mathcal{M} \in \mathbb{R}^{I_1 \times I_2, \dots, \times I_N}$ with the matrix $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{M} \times_n \mathbf{B}$, is an N-dimensional tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times I_2, \dots, \times J_n, \dots, \times I_N}$, whose elements are computed by

$$c_{i_1 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} m_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} b_{j_n i_n}. \quad (1)$$

The n-mode product $\mathcal{M} \times_n \mathbf{B}$ can also be computed by matrix multiplication $\mathbf{C}_{(n)} = \mathbf{B} \mathbf{M}_{(n)}$. For distinct modes in a series of multiplications, the order of the multiplications is irrelevant, which is

$$\mathcal{M} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n \mathbf{B} \times_m \mathbf{A} (n \neq m). \quad (2)$$

If the modes of multiplications are the same, (2) can be transformed into

$$\mathcal{M} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n (\mathbf{B} \mathbf{A}). \quad (3)$$

Given the collection of matrices $\mathbf{D}_n \in \mathbb{R}^{J_n \times I_n} (n = 1, 2, \dots, N)$, we define the tensor $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2, \dots, \times J_N}$ as

$$\mathcal{G} = \mathcal{M} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \dots \times_N \mathbf{D}_N. \quad (4)$$

Then, we have

$$\mathbf{g} = (\mathbf{D}_N \otimes \mathbf{D}_{N-1} \otimes \dots \otimes \mathbf{D}_1) \mathbf{m}, \quad (5)$$

where $\mathbf{g} = \text{vec}(\mathcal{G}) \in \mathbb{R}^J (J = \prod_{n=1}^N J_n)$, and $\mathbf{m} = \text{vec}(\mathcal{M}) \in \mathbb{R}^I (I = \prod_{n=1}^N I_n)$ are vectors obtained by stacking all the 1-mode vectors of the tensors \mathcal{C} and \mathcal{M} , respectively, and the symbol \otimes denotes Kronecker product.

Given the tensor \mathcal{M} , $\|\mathcal{M}\|_0$ denotes the ℓ_0 norm (i.e., the number of non-zero elements of \mathcal{M}), $\|\mathcal{M}\|_1 = \sum_{i_1, \dots, i_N} |m_{i_1 \dots i_N}|$ is the ℓ_1 norm, and $\|\mathcal{M}\|_F = \sqrt{\sum_{i_1, \dots, i_N} |m_{i_1 \dots i_N}|^2}$ is Frobenius norm.

III. PROBLEM FORMULATION

In this paper, both HSIs and MSIs are denoted as three-dimensional tensors indexed by three exploratory variables (w, h, s) , where w , h , and s are the indexes of the width, height, and spectral modes, respectively.

The target HR-HSI is denoted by a three-dimensional tensor $\mathcal{X} \in \mathbb{R}^{W \times H \times S}$, where W , H and S are the dimensions of the width (first) mode, height (second) mode and spectral (third) mode, respectively. The tensor $\mathcal{Y} \in \mathbb{R}^{w \times h \times S}$ denotes the acquired LR-HSI, which is spatially downsampled with respect to (w.r.t.) \mathcal{X} , that is, $W > w$ and $H > h$. The tensor $\mathcal{Z} \in \mathbb{R}^{W \times H \times s}$ denotes the HR-MSI of the same scene, which is spectrally downsampled w.r.t. \mathcal{X} , that is, $S > s$. The goal of fusion is to estimate \mathcal{X} from the observations \mathcal{Y} and \mathcal{Z} .

A. Matrix Factorization based Fusion

The matrix factorization based fusion methods assume that each spectral vector of the target HR-HSI can be written as a linear combination of a small number of distinct spectral signatures [29]. Formally, we have

$$\mathbf{X}_{(3)} = \mathbf{D}\mathbf{A}, \quad (6)$$

where $\mathbf{X}_{(3)} \in \mathbb{R}^{S \times WH}$ is the 3-mode (spectral-mode) unfolding matrix of the tensor \mathcal{X} , and the matrices $\mathbf{D} \in \mathbb{R}^{S \times L}$ and $\mathbf{A} \in \mathbb{R}^{L \times WH}$ denote, respectively, a spectral basis and the corresponding coefficient matrix. Both the LR-HSI and HR-MSI can be modeled as downsampled versions of the HR-HSI [26], [28], i.e.,

$$\mathbf{Y}_{(3)} = \mathbf{X}_{(3)}\mathbf{M}, \quad \mathbf{Z}_{(3)} = \mathbf{P}_3\mathbf{X}_{(3)}, \quad (7)$$

where $\mathbf{Y}_{(3)} \in \mathbb{R}^{S \times wh}$ and $\mathbf{Z}_{(3)} \in \mathbb{R}^{s \times WH}$ are the spectral mode (3-mode) unfolding matrices of \mathcal{Y} and \mathcal{Z} , respectively. $\mathbf{M} \in \mathbb{R}^{WH \times wh}$ is a matrix modelling the point spread function (PSF) and the spatial subsampling process in the hyperspectral sensor. $\mathbf{P}_3 \in \mathbb{R}^{s \times S}$ is a matrix modelling spectral downsampling in the multispectral sensor, whose rows contain the spectral response of the multispectral sensor. In the matrix factorization based fusion approaches, the goal is to estimate the spectral basis \mathbf{D} and coefficient matrix \mathbf{A} from $\mathbf{Y}_{(3)}$ and $\mathbf{Z}_{(3)}$.

B. Tensor Factorization based Fusion

In contrast with the matrix factorization based methods, the proposed approach assumes that the HR-HSI can be represented as a core tensor multiplied by the dictionaries of the width, height, and spectral modes. This tensor model is illustrated in Fig. 1. Formally, we represent the HR-HSI as

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}, \quad (8)$$

where the matrices $\mathbf{W} \in \mathbb{R}^{W \times n_w}$, $\mathbf{H} \in \mathbb{R}^{H \times n_h}$, and $\mathbf{S} \in \mathbb{R}^{S \times n_s}$ denote the dictionaries of the width mode with n_w atoms, height mode with n_h atoms, and spectral mode with n_s atoms, respectively. The tensor $\mathcal{C} \in \mathbb{R}^{n_w \times n_h \times n_s}$ holds the coefficient of \mathcal{X} over the three dictionaries. In this formulation, the information of three modes is incorporated into an unified model. HR-HSIs of the real world have two

important characteristics (see, e.g., [40]): 1) HR-HSIs spectral vectors are known to live, with very good approximation, in a low dimensional subspaces, and 2) HR-HSIs are spatially self-similar. The low dimensionality in the spectral domain means the it possible to find a spectral dictionary \mathbf{S} having a small number of atoms n_s ; the spatial self-similarities means the it possible to find a spatial dictionaries \mathbf{W} and \mathbf{H} , with n_w and n_h atoms, allowing a sparse representations of the respective spatial modes. The low dimensionality in the spectral domain and sparse spatial representations may be jointly pursued by promoting low dimensional core tensors.

The acquired LR-HSI \mathcal{Y} is a spatially downsampled version of the LR-HSI \mathcal{X} . Assuming that the point spread function (PSF) of the hyperspectral sensor and the downsampling matrices of the width mode and height modes are separable, then we may write

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{P}_1 \times_2 \mathbf{P}_2, \quad (9)$$

where $\mathbf{P}_1 \in \mathbb{R}^{w \times W}$ and $\mathbf{P}_2 \in \mathbb{R}^{h \times H}$ are the downsampling matrices along the width and height modes, respectively, which describe the spatial response of the imaging sensors. Under the separability assumption and substituting (8) into (9), the LR-HSI \mathcal{Y} can be written as

$$\begin{aligned} \mathcal{Y} &= \mathcal{C} \times_1 (\mathbf{P}_1 \mathbf{W}) \times_2 (\mathbf{P}_2 \mathbf{H}) \times_3 \mathbf{S} \\ &= \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S}, \end{aligned} \quad (10)$$

where $\mathbf{W}^* = \mathbf{P}_1 \mathbf{W} \in \mathbb{R}^{w \times n_w}$ and $\mathbf{H}^* = \mathbf{P}_2 \mathbf{H} \in \mathbb{R}^{h \times n_h}$ are the downsampled dictionaries of the width and height modes, respectively. Regarding the PSF, the separability assumption is valid, for example, for boxcar and Gaussian convolution kernels with major axis aligned with the spatial unit vectors. The assumption of separable sensing operators brings great advantages in tensorial calculus, as discussed in [41]. Regarding subsampling, the separability assumption means that the action of the spatial subsampling matrix \mathbf{M} is decoupled with respect to the two spatial modes of \mathcal{X} , and therefore we have

$$\mathbf{M} = (\mathbf{P}_2 \otimes \mathbf{P}_1)^T. \quad (11)$$

In this way, the spatial downsampling processes in (2) and (7) are equivalent.

The HR-MSI \mathcal{Z} is a spectrally downsampled version of \mathcal{X} , i.e.,

$$\mathcal{Z} = \mathcal{X} \times_3 \mathbf{P}_3, \quad (12)$$

where $\mathbf{P}_3 \in \mathbb{R}^{s \times S}$ is the downsampling matrix of the spectral mode. Substituting (8) into (12), the HR-MSI \mathcal{Z} can also be represented as

$$\begin{aligned} \mathcal{Z} &= \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 (\mathbf{P}_3 \mathbf{S}) \\ &= \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^*, \end{aligned} \quad (13)$$

where $\mathbf{S}^* = \mathbf{P}_3 \mathbf{S} \in \mathbb{R}^{s \times n_s}$ is the downsampled spectral dictionary.

To reconstruct the HR-HSI, we need to estimate the dictionaries \mathbf{W} , \mathbf{H} , \mathbf{S} and corresponding core tensor \mathcal{C} , as illustrated in Fig. 1.

IV. PROPOSED CSTF APPROACH

Since LR-HSI and HR-MSI are spatially and spectrally downsampled, respectively, the fusion problem is severely ill-posed. Therefore, we need to use some form prior information of the unknown HR-HSI to regularize the fusion problem. Spectral sparsity has been widely used to regularize various ill-posed inverse problems in the context of HSI processing [42]–[45]. This form of regularization assumes each spectral vector of the HSI (i.e., the 3-mode unfolding vectors of \mathcal{X}) can be represented as the linear combinations of a small number of distinct spectral signatures. These methods only take the sparsity of the spectral mode into consideration. In this paper, we extend the sparsity regularization to the spatial domain, linked with the HSI self-similarity, by exploiting the tensor factorization (8). Specifically, we do not only consider the sparsity in the spectral mode but also in the width and height modes. In the tensor factorization, we claim that the HR-HSI admits a sparse representation over the three dictionaries, i.e., the core tensor \mathcal{C} is sparse.

Taking (10) and (13) into account, we formulate the fusion problem as the following constrained least-squares optimization problem:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}} & \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 \\ & + \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2, \\ \text{s.t. } & \| \mathcal{C} \|_0 \leq N, \end{aligned} \quad (14)$$

where $\|\cdot\|_F$ denotes the the Frobenius norm, and N is the maximum number of non-zero elements in \mathcal{C} . Optimization (14) is nonconvex due to the constraint expressed in terms of the ℓ_0 norm. In order to obtain a treatable optimization, we replace the ℓ_0 norm with the ℓ_1 norm and formulate the unconstrained version

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}} & \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 \\ & + \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \lambda \| \mathcal{C} \|_1, \end{aligned} \quad (15)$$

where λ is a sparsity regularization parameter. Problem (15) is still non-convex, and the solutions for \mathbf{W} , \mathbf{H} , \mathbf{S} , and \mathcal{C} are not unique. However, the objective function in (15) is convex for each block of variables, keeping with other variables fixed. We use the proximal alternating optimization (PAO) scheme (see, e.g. [46], [47]) to solve (15), which can be guaranteed to converge to a critical point under certain conditions. Specifically, \mathbf{W} , \mathbf{H} , \mathbf{S} , and \mathcal{C} , are iteratively updated as follows:

$$\begin{aligned} \mathbf{W} &= \underset{\mathbf{W}}{\operatorname{argmin}} f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) + \beta \| \mathbf{W} - \mathbf{W}_{\text{pre}} \|_F^2 \\ \mathbf{H} &= \underset{\mathbf{H}}{\operatorname{argmin}} f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) + \beta \| \mathbf{H} - \mathbf{H}_{\text{pre}} \|_F^2 \\ \mathbf{S} &= \underset{\mathbf{S}}{\operatorname{argmin}} f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) + \beta \| \mathbf{S} - \mathbf{S}_{\text{pre}} \|_F^2 \\ \mathcal{C} &= \underset{\mathcal{C}}{\operatorname{argmin}} f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) + \beta \| \mathcal{C} - \mathcal{C}_{\text{pre}} \|_F^2, \end{aligned} \quad (16)$$

where objective function $f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C})$ is implicitly defined in (15), β is a positive number, and $(\cdot)_{\text{pre}}$ represents the estimated blocks of variables in the previous iteration. Next, we present the optimizations for \mathbf{W} , \mathbf{H} , \mathbf{S} , and \mathcal{C} in detail.

A. Optimization with respect to \mathbf{W}

With \mathbf{H} , \mathbf{S} , and \mathcal{C} fixed, the optimization with respect to (w.r.t.) \mathbf{W} in (16) can be written as

$$\begin{aligned} \min_{\mathbf{W}} & \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 + \\ & \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \\ & \beta \| \mathbf{W} - \mathbf{W}_{\text{pre}} \|_F^2, \end{aligned} \quad (17)$$

where \mathbf{W}_{pre} represents the estimated dictionary of width mode in the previous iteration. Using the properties of n -mode matrix unfolding, problem (17) may be rewritten as

$$\begin{aligned} \min_{\mathbf{W}} & \| \mathbf{Y}_{(1)} - \mathbf{P}_1 \mathbf{W} \mathbf{A}_w \|_F^2 + \| \mathbf{Z}_{(1)} - \mathbf{W} \mathbf{B}_w \|_F^2 + \\ & \beta \| \mathbf{W} - \mathbf{W}_{\text{pre}} \|_F^2, \end{aligned} \quad (18)$$

where $\mathbf{Y}_{(1)}$ and $\mathbf{Z}_{(1)}$ are the width-mode (1-mode) unfolding matrix of tensors \mathcal{Y} and \mathcal{Z} , respectively, $\mathbf{A}_w = (\mathcal{C} \times_2 \mathbf{H}^* \times_3 \mathbf{S})_{(1)}$, and $\mathbf{B}_w = (\mathcal{C} \times_2 \mathbf{H} \times_3 \mathbf{S}^*)_{(1)}$. Problem (18) is quadratic and its unique solution (we are assuming that $\beta > 0$) amounts to compute the general Sylvester matrix equation

$$\begin{aligned} \mathbf{P}_1^T \mathbf{P}_1 \mathbf{W} \mathbf{A}_w \mathbf{A}_w^T + \mathbf{W} (\mathbf{B}_w \mathbf{B}_w^T + \beta \mathbf{I}) = \\ \mathbf{P}_1^T \mathbf{Y}_{(1)} \mathbf{A}_w^T + \mathbf{Z}_{(1)} \mathbf{B}_w^T + \beta \mathbf{W}_{\text{pre}}. \end{aligned} \quad (19)$$

We use conjugate gradient (CG) [48] to solve (19). Regarding the application of CG, we make three remarks: a) the system matrix associated with (19) is symmetric and positive definite, which is a necessary and sufficient condition to directly apply CG; b) the step of the heaviest computation in applying CG is the multiplication of the system matrix times a vector, which can be carried out very efficiently in the matrix representation, and finally, c) CG converges in only a few iterations. In fact, the convergence of CG is linear and depends on the ratio between the largest and the smallest eigenvalues of the system matrix associated with (19) [49]. In our experiments, we have systematically observed that 30 iterations yield a very good approximation of the solution of (19).

B. Optimization with respect to \mathbf{H}

With \mathbf{W} , \mathbf{S} , and \mathcal{C} fixed, the optimization w.r.t \mathbf{H} in (16) can be written as

$$\begin{aligned} \min_{\mathbf{H}} & \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 + \\ & \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \\ & \beta \| \mathbf{H} - \mathbf{H}_{\text{pre}} \|_F^2, \end{aligned} \quad (20)$$

where \mathbf{H}_{pre} represents the estimated dictionary of height mode in the previous iteration. Similarly to the estimation of \mathbf{W} , problem (20) may be rewritten as

$$\begin{aligned} \min_{\mathbf{H}} & \| \mathbf{Y}_{(2)} - \mathbf{P}_2 \mathbf{H} \mathbf{A}_h \|_F^2 + \| \mathbf{Z}_{(2)} - \mathbf{H} \mathbf{B}_h \|_F^2 + \\ & \beta \| \mathbf{H} - \mathbf{H}_{\text{pre}} \|_F^2, \end{aligned} \quad (21)$$

where $\mathbf{Y}_{(2)}$ and $\mathbf{Z}_{(2)}$ are the width-mode (2-mode) unfolding matrix of tensors \mathcal{Y} and \mathcal{Z} , respectively, $\mathbf{A}_h = (\mathcal{C} \times_1 \mathbf{W}^* \times_3 \mathbf{S})_{(2)}$, and $\mathbf{B}_h = (\mathcal{C} \times_1 \mathbf{W} \times_3 \mathbf{S}^*)_{(2)}$. Problem (21) is quadratic

and its unique solution (we are assuming that $\beta > 0$) amounts to compute the general Sylvester matrix equation

$$\begin{aligned} \mathbf{P}_2^T \mathbf{P}_2 \mathbf{W} \mathbf{A}_h \mathbf{A}_h^T + \mathbf{W} (\mathbf{B}_h \mathbf{B}_h^T + \beta \mathbf{I}) = \\ \mathbf{P}_2^T \mathbf{Y}_{(2)} \mathbf{A}_h^T + \mathbf{Z}_{(2)} \mathbf{B}_h^T + \beta \mathbf{H}_{\text{pre}}. \end{aligned} \quad (22)$$

We use conjugate gradient (CG) to solve (22); the convergence is linear and achieved just in a few iterations. The justification is as in the optimization w.r.t. \mathbf{W} .

C. Optimization with respect to \mathbf{S}

With \mathbf{W} , \mathbf{H} , and \mathcal{C} fixed, the optimization w.r.t \mathbf{S} in (16) can be written as

$$\begin{aligned} \min_{\mathbf{S}} \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 + \\ \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \\ \beta \| \mathbf{S} - \mathbf{S}_{\text{pre}} \|_F^2, \end{aligned} \quad (23)$$

where \mathbf{S}_{pre} represents the estimated spectral dictionary in the previous iteration. Proceeding as in the two previous subsections, we may write

$$\begin{aligned} \min_{\mathbf{S}} \| \mathbf{Y}_{(3)} - \mathbf{S} \mathbf{A}_s \|_F^2 + \| \mathbf{Z}_{(3)} - \mathbf{P}_3 \mathbf{S} \mathbf{B}_s \|_F^2 + \\ \beta \| \mathbf{S} - \mathbf{S}_{\text{pre}} \|_F^2, \end{aligned} \quad (24)$$

where $\mathbf{Y}_{(3)}$ and $\mathbf{Z}_{(3)}$ are the spectral-mode (3-mode) unfolding matrix of tensors \mathcal{Y} and \mathcal{Z} , respectively, $\mathbf{A}_s = (\mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^*)_{(3)}$, and $\mathbf{B}_s = (\mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H})_{(3)}$. Problem (24) is quadratic and its unique solution (we are assuming that $\beta > 0$) amounts to compute the general Sylvester matrix equation

$$\begin{aligned} \mathbf{S} (\mathbf{A}_s \mathbf{A}_s^T + \beta \mathbf{I}) + \mathbf{P}_3^T \mathbf{P}_3 \mathbf{W} \mathbf{B}_s \mathbf{B}_s^T = \\ \mathbf{Y}_{(3)} \mathbf{A}_s^T + \mathbf{P}_3^T \mathbf{Z}_{(3)} \mathbf{B}_s^T + \beta \mathbf{S}_{\text{pre}}. \end{aligned} \quad (25)$$

We use conjugate gradient (CG) to solve (25); the convergence is linear and achieved just in a few iterations. The justification is as in the optimization w.r.t. \mathbf{W} .

D. Optimization with respect to \mathcal{C}

With the dictionaries \mathbf{W} , \mathbf{H} , and \mathbf{S} fixed, the optimization w.r.t. the core tensor \mathcal{C} can be formulated as

$$\begin{aligned} \min_{\mathcal{C}} \| \mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 + \lambda \| \mathcal{C} \|_1 + \\ \| \mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \beta \| \mathcal{C} - \mathcal{C}_{\text{pre}} \|_F^2, \end{aligned} \quad (26)$$

where \mathcal{C}_{pre} represents the estimated core tensor in the previous iteration. Problem (26) is convex and can be solved efficiently by the alternating direction method of multipliers (ADMM) [50], [51]. By introducing the splitting variables $\mathcal{C}_1 = \mathcal{C}$ and $\mathcal{C}_2 = \mathcal{C}$, we may write the unconstrained optimization (26) as the equivalent constrained form

$$\begin{aligned} \min_{\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2} g(\mathcal{C}) + g_1(\mathcal{C}_1) + g_2(\mathcal{C}_2) \\ \text{s.t.: } \mathcal{C}_1 = \mathcal{C}, \mathcal{C}_2 = \mathcal{C} \end{aligned} \quad (27)$$

where

$$g(\mathcal{C}) = \lambda \| \mathcal{C} \|_1 + \beta \| \mathcal{C} - \mathcal{C}_{\text{pre}} \|_F^2 \quad (28)$$

$$g_1(\mathcal{C}_1) = \| \mathcal{Y} - \mathcal{C}_1 \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 \quad (29)$$

$$g_2(\mathcal{C}_2) = \| \mathcal{Z} - \mathcal{C}_2 \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2. \quad (30)$$

The optimization template (27) corresponds to the standard ADMM, provided that the optimization variables \mathcal{C}_1 and \mathcal{C}_2 are considered as only one block. The augmented Lagrangian function for (27) is

$$\begin{aligned} L(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2, \mathcal{V}_1, \mathcal{V}_2) = \lambda \| \mathcal{C} \|_1 + \beta \| \mathcal{C} - \mathcal{C}_{\text{pre}} \|_F^2 + \\ \| \mathcal{Y} - \mathcal{C}_1 \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2 + \mu \| \mathcal{C} - \mathcal{C}_1 - \mathcal{V}_1 \|_F^2 + \\ \| \mathcal{Z} - \mathcal{C}_2 \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^* \|_F^2 + \mu \| \mathcal{C} - \mathcal{C}_2 - \mathcal{V}_2 \|_F^2, \end{aligned} \quad (31)$$

where \mathcal{V}_1 and \mathcal{V}_2 are the scaled Lagrangian multipliers, and μ is a penalty parameter. ADMM iterations are as follows:

for $t = 1, \dots$

$$\mathcal{C} \leftarrow \arg \min_{\mathcal{C}} L(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2, \mathcal{V}_1, \mathcal{V}_2) \quad (32)$$

$$\mathcal{C}_1 \leftarrow \arg \min_{\mathcal{C}_1} L(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2, \mathcal{V}_1, \mathcal{V}_2) \quad (33)$$

$$\mathcal{C}_2 \leftarrow \arg \min_{\mathcal{C}_2} L(\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2, \mathcal{V}_1, \mathcal{V}_2) \quad (34)$$

$$\mathcal{V}_1 \leftarrow \mathcal{V}_1 - (\mathcal{C} - \mathcal{C}_1) \quad (35)$$

$$\mathcal{V}_2 \leftarrow \mathcal{V}_2 - (\mathcal{C} - \mathcal{C}_2) \quad (36)$$

end

We remark that the optimizations w.r.t \mathcal{C}_1 and \mathcal{C}_2 are independent because L is decoupled w.r.t. these variables. Below, we detail the optimizations (32), (33), and (34).

1) *Update \mathcal{C} :* From (31), we have

$$\mathcal{C} \in \operatorname{argmin}_{\mathcal{C}} \lambda \| \mathcal{C} \|_1 + \beta \| \mathcal{C} - \mathcal{C}_{\text{pre}} \|_F^2 + \mu \| \mathcal{C} - \mathcal{C}_1 - \mathcal{V}_1 \|_F^2 + \mu \| \mathcal{C} - \mathcal{C}_2 - \mathcal{V}_2 \|_F^2, \quad (37)$$

$$+ \mu \| \mathcal{C} - \mathcal{C}_1 - \mathcal{V}_1 \|_F^2 + \mu \| \mathcal{C} - \mathcal{C}_2 - \mathcal{V}_2 \|_F^2, \quad (38)$$

which yields the closed-form solution

$$\mathcal{C} = \operatorname{soft} \left[\frac{\mu(\mathcal{C}_1 + \mathcal{V}_1 + \mathcal{C}_2 + \mathcal{V}_2) + \beta \mathcal{C}_{\text{pre}}}{2\mu + \beta}, \frac{\lambda}{4\mu + 2\beta} \right], \quad (39)$$

where $\operatorname{soft}(a,b) = \operatorname{sign}(a) * \max(|a| - b, 0)$.

2) *Update \mathcal{C}_1 :* From (31), we have

$$\begin{aligned} \mathcal{C}_1 \in \operatorname{argmin}_{\mathcal{C}_1} \mu \| \mathcal{C}_1 - \mathcal{C} + \mathcal{V}_1 \|_F^2 + \\ \| \mathcal{Y} - \mathcal{C}_1 \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S} \|_F^2. \end{aligned} \quad (40)$$

Based on (4) and (5), problem (40) is equivalent to

$$\operatorname{argmin}_{\mathbf{c}_1} \mu \| \mathbf{c}_1 - \mathbf{c} + \mathbf{v}_1 \|_F^2 + \| \mathbf{y} - \mathbf{D}_1 \mathbf{c}_1 \|_F^2, \quad (41)$$

where the vectors $\mathbf{c}_1 = \operatorname{vec}(\mathcal{C}_1)$, $\mathbf{v}_1 = \operatorname{vec}(\mathcal{V}_1)$, $\mathbf{c} = \operatorname{vec}(\mathcal{C})$, and $\mathbf{y} = \operatorname{vec}(\mathcal{Y})$ are obtained by vectorizing the tensors \mathcal{C}_1 , \mathcal{V}_1 , \mathcal{C} , and \mathcal{Y} , respectively, and

$$\mathbf{D}_1 = \mathbf{S} \otimes \mathbf{H}^* \otimes \mathbf{W}^*. \quad (42)$$

Problem (41) has the closed-form solution

$$\mathbf{c}_1 = (\mathbf{D}_1^T \mathbf{D}_1 + \mu \mathbf{I})^{-1} (\mathbf{D}_1^T \mathbf{y} + \mu \mathbf{c} - \mu \mathbf{v}_1). \quad (43)$$

Dictionary $\mathbf{D}_1 \in \mathbb{R}^{whS \times n_w n_h n_s}$ is very large making it very hard to solve the system defined by (43). We note however that

$$\begin{aligned} (\mathbf{D}_1^T \mathbf{D}_1 + \mu \mathbf{I})^{-1} = & (\mathbf{P}_3 \otimes \mathbf{P}_2 \otimes \mathbf{P}_1) \\ & (\mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 + \mu \mathbf{I})^{-1} \\ & (\mathbf{P}_3^T \otimes \mathbf{P}_2^T \otimes \mathbf{P}_1^T), \end{aligned} \quad (44)$$

where \mathbf{P}_i and Σ_i , for $i = 1, 2, 3$, are unitary matrices and non-negative diagonal matrices holding the eigenvectors and eigenvalues of, respectively, $\mathbf{W}^{*T}\mathbf{W}^*$, $\mathbf{H}^{*T}\mathbf{H}^*$, and $\mathbf{S}^T\mathbf{S}$. Therefore, $(\Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 + \mu\mathbf{I})^{-1}$ is diagonal and very easy to compute. Finally, to compute \mathbf{c}_1 , we note that the actions of \mathbf{P}_i^T and of \mathbf{P}_i are just i -mode products and the action of multiplication $(\Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 + \mu\mathbf{I})^{-1}$ is elementwise multiplication. Besides, the term $\mathbf{D}_1^T\mathbf{y}$ in (43) can be computed via the following equation

$$\mathbf{D}_1^T\mathbf{y} = \text{vec} \left(\mathcal{Y} \times {}_1\mathbf{W}^{*T} \times {}_2\mathbf{H}^{*T} \times {}_3\mathbf{S}^T \right), \quad (45)$$

where $\text{vec}(\cdot)$ is the vectorization operation.

3) *Update \mathcal{C}_2 :* From (31), we have

$$\begin{aligned} \mathcal{C}_2 = \underset{\mathcal{C}_2}{\operatorname{argmin}} \mu & \| \mathcal{C}_2 - \mathcal{C} + \mathcal{V}_2 \|_F^2 \\ & + \| \mathcal{Z} - \mathcal{C}_2 \times {}_1\mathbf{W} \times {}_2\mathbf{H} \times {}_3\mathbf{S}^* \|_F^2. \end{aligned} \quad (46)$$

Based on the equivalence between (4) and (5), problem (46) is equivalent to

$$\underset{\mathbf{c}_2}{\operatorname{argmin}} \mu \| \mathbf{c}_2 - \mathbf{c} + \mathbf{v}_2 \|_F^2 + \| \mathbf{z} - \mathbf{D}_2\mathbf{c}_2 \|_F^2, \quad (47)$$

where vectors $\mathbf{c}_2 = \text{vec}(\mathcal{C}_2)$, $\mathbf{v}_2 = \text{vec}(\mathcal{V}_2)$, $\mathbf{c} = \text{vec}(\mathcal{C})$, and $\mathbf{z} = \text{vec}(\mathcal{Z})$ are obtained vectorizing the tensors \mathcal{C}_2 , \mathcal{V}_2 , \mathcal{C} , and \mathcal{Z} , respectively, and

$$\mathbf{D}_2 = \mathbf{S}^* \otimes \mathbf{H} \otimes \mathbf{W}. \quad (48)$$

Problem (47) has the closed-form solution

$$\mathbf{c}_2 = (\mathbf{D}_2^T \mathbf{D}_2 + \mu\mathbf{I})^{-1} (\mathbf{D}_2^T \mathbf{y} + \mu\mathbf{c} - \mu\mathbf{v}_2). \quad (49)$$

Dictionary $\mathbf{D}_2 \in \mathbb{R}^{WHs \times n_w n_h n_s}$ is vary large making it very hard to solve the system defined in (49). We note however that

$$\begin{aligned} (\mathbf{D}_2^T \mathbf{D}_2 + \mu\mathbf{I})^{-1} &= (\tilde{\mathbf{P}}_3 \otimes \tilde{\mathbf{P}}_2 \otimes \tilde{\mathbf{P}}_1) \\ &\quad (\tilde{\Sigma}_3 \otimes \tilde{\Sigma}_2 \otimes \tilde{\Sigma}_1 + \mu\mathbf{I})^{-1} \\ &\quad (\tilde{\mathbf{P}}_3^T \otimes \tilde{\mathbf{P}}_2^T \otimes \tilde{\mathbf{P}}_1^T), \end{aligned} \quad (50)$$

where $\tilde{\mathbf{P}}_i$ and $\tilde{\Sigma}_i$, for $i = 1, 2, 3$, are unitary matrices and non-negative diagonal matrices holding the eigenvectors and eigenvalues of, respectively, $\mathbf{W}^T\mathbf{W}$, $\mathbf{H}^T\mathbf{H}$, and $\mathbf{S}^{*T}\mathbf{S}^*$. Therefore, $(\tilde{\Sigma}_3 \otimes \tilde{\Sigma}_2 \otimes \tilde{\Sigma}_1 + \mu\mathbf{I})^{-1}$ is diagonal and very easy to compute. Finally, to compute \mathbf{c}_2 , we note that the actions of $\tilde{\mathbf{P}}_i^T$ and of $\tilde{\mathbf{P}}_i$ are just i -mode products and the action of $(\tilde{\Sigma}_3 \otimes \tilde{\Sigma}_2 \otimes \tilde{\Sigma}_1 + \mu\mathbf{I})^{-1}$ is elementwise multiplication.

4) *Update of the Lagrangian multipliers \mathcal{V}_1 and \mathcal{V}_2 :* From (35)-(36), we have

$$\begin{aligned} \mathcal{V}_1 &\leftarrow \mathcal{V}_1 - (\mathcal{C} - \mathcal{C}_1) \\ \mathcal{V}_2 &\leftarrow \mathcal{V}_2 - (\mathcal{C} - \mathcal{C}_2). \end{aligned} \quad (51)$$

E. Convergence of the ADMM Algorithm to \mathcal{C}

Algorithm 1 summarizes the estimation of core tensor \mathcal{C} . The ADMM steps (39), (43), (49), and (51) are iterated until convergence.

The following proposition states the convergence of the ADMM **Algorithm 1**.

Algorithm 1: Solve optimization (26) with regard to core tensor \mathcal{C} via the ADMM

Input: $\mathcal{Y}, \mathcal{Z}, \mathbf{W}, \mathbf{W}^*, \mathbf{H}, \mathbf{H}^*, \mathbf{S}, \mathbf{S}^*, \mathcal{C}_{\text{pre}}, \mu > 0$, $\beta > 0$, and $\lambda > 0$.

Output: \mathcal{C}

while not converged **do**

Step 1 Update core tensor \mathcal{C} via equation (39)

Step 2 Update variable \mathcal{C}_1 via (43)

Step 3 Update variable \mathcal{C}_2 via (49)

Step 4 Update Lagrangian multipliers \mathcal{V}_1 and \mathcal{V}_2 via (51)

end while

Proposition 1. *The ADMM Algorithm 1 converges to a solution of (26) for any $\mu > 0$, $\lambda \geq 0$, and $\beta \geq 0$. If $\beta > 0$, the solution is unique.*

Proof: The four terms in (26) are closed, proper, and convex functions with domain $\mathbb{R}^{n_w n_h n_s}$. Therefore, the functions g , defined in (28), and $g_1 + g_2$, defined in (29) and (30), are also closed, proper, and convex. The null space of the linear mapping

$$\begin{aligned} G : \mathbb{R}^{n_w n_h n_s} &\rightarrow \mathbb{R}^{2n_w n_h n_s} \\ \mathcal{C} &\mapsto (\mathcal{C}, \mathcal{C}) \end{aligned}$$

is just the zero vector. If $\beta = \lambda = 0$, the optimization (26) has solutions as it is the sum of semi-positive quadratic terms; if $\lambda > 0$, the objective function is coercive and then the optimization has solutions; if $\beta > 0$, the objective function is strictly convex and then the solution is unique. Therefore, if $\mu > 0$, we may invoke the theorem by Eckstein and Bertsekas [50, Theorem 8], which, asserts that the sequence of tensors \mathcal{C} obtained by (39) converges to a solution of (26). ■

As stopping criterion for **Algorithm 1**, and following [51, Ch. 3], we use the termination criterion $\|r^k\|_F \leq \text{tol}_1$ and $\|s^k\|_F \leq \text{tol}_1$, where r^k and s^k are the so-called primal and dual residuals, respectively. We set $\text{tol}_1 = 0.04$ in the experiments reported in Section V. In practice, we have observed that 40 iterations are enough to satisfy the termination criterion.

F. CSTF based Fusion

Optimization problem (15) is non-convex. Therefore, the initialization should be carefully designed to avoid poor local minima. Since the spectral information is mostly in the LR-HSI and the spatial information is mostly in the HR-MSI, we initialize the spectral dictionary \mathbf{S} based on the former and the dictionaries of the width mode \mathbf{W} and height mode \mathbf{H} based on the latter. The spectral dictionary \mathbf{S} is initialized via simplex identification split augmented Lagrangian (SISAL) [52] method, which efficiently identifies a minimum volume simplex containing the LR-HSI spectral vectors. The dictionaries \mathbf{W} and \mathbf{H} are initialized from the HR-MSI unfolding matrices ($\mathbf{Z}_{(1)}$ for the width-mode and $\mathbf{Z}_{(2)}$ for the height-mode) via dictionary-updates-cycles KSVD (DUC-KSVD) [53], which promotes sparse representations.

Algorithm 2 summarizes the proposed CSTF method for the HSI-MSI fusion, which aims at solving problem (15); it implements the PAO scheme shown in (16). To set the stage

Algorithm 2: CSTF-based HSI-MSI Fusion

Input: $\mathcal{Y}, \mathcal{Z}, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \beta > 0$, and $\lambda > 0$
Output: HR-HSI \mathcal{X}

Initialize \mathbf{W} and \mathbf{H} with DUC-KSVD algorithm

Initialize \mathbf{S} with SISAL algorithms

Initialize \mathcal{C} with **Algorithm 1**

while not converged **do**

step 1: Update dictionary \mathbf{W} by solving (19) with CG

$$\mathbf{W}^* = \mathbf{P}_1 \mathbf{W}, \mathbf{W}_{\text{pre}} = \mathbf{W}$$

step 2: Update dictionary \mathbf{H} via (22) with CG

$$\mathbf{H}^* = \mathbf{P}_2 \mathbf{H}, \mathbf{H}_{\text{pre}} = \mathbf{H}$$

step 3: Update dictionary \mathbf{S} via (25) with CG

$$\mathbf{S}^* = \mathbf{P}_3 \mathbf{S}, \mathbf{S}_{\text{pre}} = \mathbf{S}$$

Step 4 Update core tensor \mathcal{C} via **Algorithm 1**

$$\mathcal{C}_{\text{pre}} = \mathcal{C}$$

end while

Estimate the HR-HSI \mathcal{X} via the equation (8)

for proving the convergence of **Algorithm 2**, it is convenient to rewrite the function f as

$$f(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) = Q(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) + f_C(\mathcal{C}), \quad (52)$$

where

$$Q(\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathcal{C}) = \|\mathcal{Y} - \mathcal{C} \times_1 \mathbf{W}^* \times_2 \mathbf{H}^* \times_3 \mathbf{S}\|_F^2 + \quad (53)$$

$$\|\mathcal{Z} - \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}^*\|_F^2$$

$$f_C(\mathcal{C}) = \lambda \|\mathcal{C}\|_1. \quad (54)$$

Proposition 2. Assume that $(\mathbf{S}^k, \mathbf{W}^k, \mathbf{H}^k, \mathcal{C}^k)_{k \in \mathbb{N}}$, i.e., the sequence generated by **Algorithm 2**, is bounded. Then, it converges to some critical point of f .

Proof: Function Q is C^1 with Lipschitz continuous gradient (consequence of the boundedness of $(\mathbf{S}^k, \mathbf{W}^k, \mathbf{H}^k, \mathcal{C}^k)_{k \in \mathbb{N}}$), f_C is proper and lower semicontinuous, and $f : \mathbb{R}^{Wn_w} \times \mathbb{R}^{Hn_h} \times \mathbb{R}^{Sn_s} \times \mathbb{R}^{n_w n_h n_s} \rightarrow \mathbb{R}$ is Kurdyka-Łojasiewicz (see Section 2.2 in [46]) and bounded below. Under these conditions, the proof is a direct application of [46, Theorem 6.2] by noting that **Algorithm 2** is an instance of algorithm (61)-(62)-(63) shown in [46] with $\mathbf{B}_i = \beta \mathbf{I}$ with $\beta > 0$ (see Remark 6.1 in [46]). ■

As stopping criterion for **Algorithm 1**, we use the changes of \mathbf{W} , \mathbf{H} , \mathbf{S} , and \mathcal{C} in two successive iterations: $\frac{\|\mathcal{C}^{k+1} - \mathcal{C}^k\|_F}{\|\mathcal{C}^k\|_F} + \frac{\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F}{\|\mathbf{W}^k\|_F} + \frac{\|\mathbf{H}^{k+1} - \mathbf{H}^k\|_F}{\|\mathbf{H}^k\|_F} + \frac{\|\mathbf{S}^{k+1} - \mathbf{S}^k\|_F}{\|\mathbf{S}^k\|_F} \leq \text{tol}_2$. We set $\text{tol}_2 = 0.04$ in the experiments reported in Section V.

G. Computational complexity of CSTF

In the CSTF, we optimize \mathbf{W} , \mathbf{H} , \mathbf{S} , and \mathcal{C} , respectively. The optimization of dictionaries \mathbf{W} , \mathbf{H} , and \mathbf{S} amounts to solve the linear system of equation with symmetric and positive definite system matrices, which are solved by the CG algorithm. In each iteration of CG, the heaviest step is the multiplication of the system matrix times a vector, whose time complexities are $O(n_w^2 W^2)$, $O(n_h^2 H^2)$, and $O(n_s^2 S^2)$ for the optimizations w.r.t. \mathbf{W} , \mathbf{H} , and \mathbf{S} , respectively. These operations may be carried out very efficiently in the matrix representation, reducing the complexity to $O(n_w^2 W + n_w W^2)$, $O(n_h^2 H + n_h H^2)$, $O(n_s^2 S + n_s S^2)$, for the optimizations of \mathbf{W} , \mathbf{H} , and \mathbf{S} , respectively. The core tensor is optimized via

the ADMM. In each iteration of ADMM, the two heaviest computation steps, shown in (43) and (49), have time complexity of $O(n_w^3 n_h^3 n_s^3)$. If we use (44), (50) to carry out those computations and tensor i-mode products, the time complexity is reduced to $O(n_w^2 n_h n_s + n_w n_h^2 n_s + n_w n_h n_s^2)$. Therefore, the time complexity of each CSTF iteration is

$$\begin{aligned} & O(N_{CG}(n_w^2 W + n_w W^2)) + \\ & O(N_{CG}(n_h^2 H + n_h H^2)) + \\ & O(N_{CG} n_s^2 S + n_s S^2)) + \\ & O(N_{ADMM}(n_w^2 n_h n_s + n_w n_h^2 n_s + n_w n_h n_s^2)), \end{aligned}$$

where N_{CG} and N_{ADMM} denote, respectively, the number of CG and of ADMM iterations.

V. EXPERIMENTS

A. Experimental Data Sets

In this section, two data sets are employed to test the performance of the proposed method. The first dataset is University of Pavia image [54] acquired by the reflective optics system imaging spectrometer (ROSIS) optical sensor over the urban area of the University of Pavia, Italy. The image is of size $610 \times 340 \times 115$ with a spatial resolution of 1.3m and a spectral coverage ranging from $0.43\mu\text{m}$ to $0.86\mu\text{m}$. The number of bands is meanwhile reduced to 93 bands after removing the water vapor absorption bands. For reasons linked with the downsampling process, only the top left 256×256 pixels are used in the experiment. We remark however that CSTF does not impose any constraint on the size of HSIs. The Pavia University of size $256 \times 256 \times 93$ image is used as the reference image. To generate the LR-HSI of size $32 \times 32 \times 93$, the HR-HSI is downsampled by averaging the 8×8 disjoint spatial blocks. A IKONOS-like spectral reflectance response filter [18] is used to generate the four-band HR-MSI with the size of $256 \times 256 \times 4$.

The second data set is Cuprite Mine in Nevada, which was taken by the NASA's Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) [55]. The image is of size $256 \times 256 \times 224$ covering the wavelength range 400nm - 2500nm with 10nm sampling interval. Following [24], we have removed the bands 1-2, 105-115, 150-170 and 223-224 of the image because of extremely low SNR and water absorptions in those bands. The LR-HSI of size $32 \times 32 \times 188$ is produced by averaging the 8×8 disjoint spatial blocks. The HR-MSI with six bands is simulated by directly selecting the 256×256 spectral images from the ground truth image, corresponding to the wavelengths $480, 560, 660, 830, 1650$ and 2220nm . These wavelengths correspond to the visible and mid-infrared range spectral channels of USGS/NASA Landsat7 satellite [24].

B. Compared Methods

We compare the proposed method with the following four current state-of-the-art HSI-MSI fusion methods: GSOMP (generalization of simultaneous orthogonal matching pursuit), [24], HySure (subspace regularization) [26], BSR (Bayesian sparse representation) [19], and NLSTF (non-local sparse tensor factorization) [30].

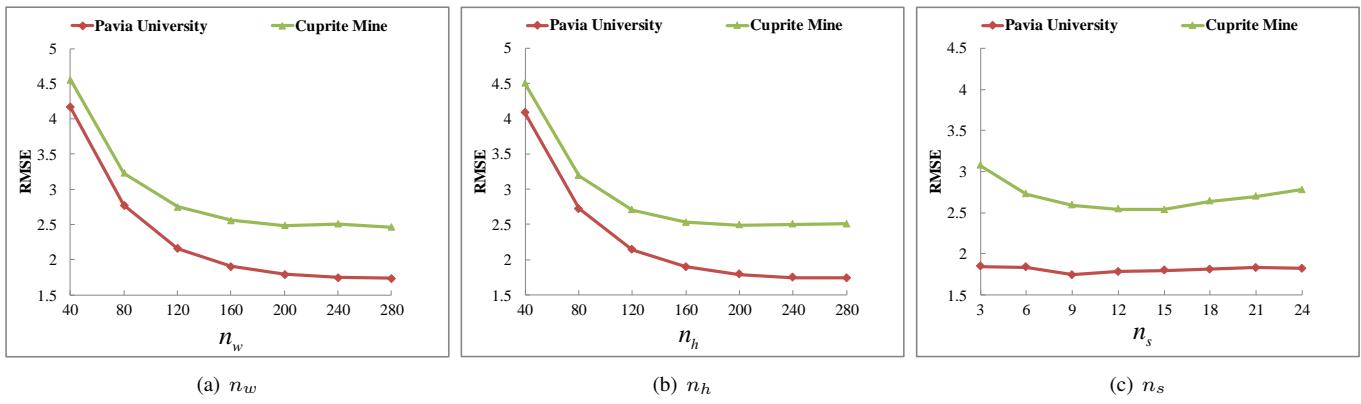


Fig. 5. The RMSE as functions of the number of atoms n_w , n_h , and n_s for the proposed CSTF method.

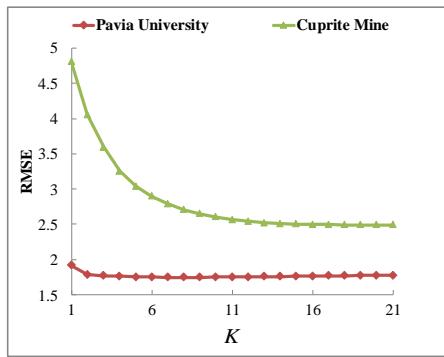


Fig. 2. The RMSE results as functions of the number of iterations K for the proposed CSTF method.

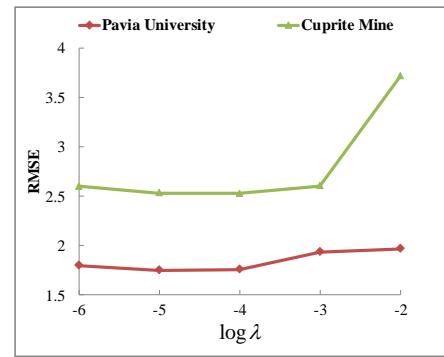


Fig. 4. The RMSE as functions of the sparsity regularization parameter $\log \lambda$ for the proposed CSTF method.

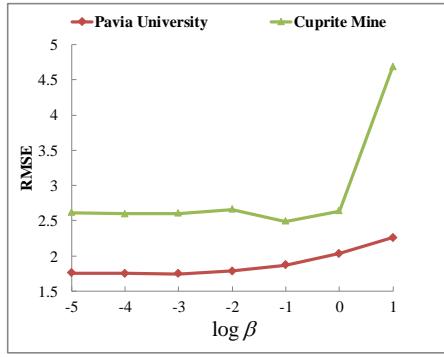


Fig. 3. The RMSE as functions of $\log \beta$ for the proposed CSTF method.

C. Quantitative Metrics

To evaluate the quality of reconstructed HSIs, six indexes are used in our study. The first index is the root mean square error (RMSE) defined as

$$\text{RMSE}(\mathcal{X}, \hat{\mathcal{X}}) = \sqrt{\frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2}{WHS}}, \quad (55)$$

where $\mathcal{X} \in \mathbb{R}^{W \times H \times S}$ and $\hat{\mathcal{X}} \in \mathbb{R}^{W \times H \times S}$ are ground truth and estimated HSIs respectively, both of them scaled to the range [0, 255]. RMSE is a measure of the estimation error. The smaller the RMSE, the better the fusion quality.

The second index is the spectral angle mapper (SAM), which is defined as the angle between the estimated pixel $\hat{\mathbf{x}}_j$ and the ground truth pixel \mathbf{x}_j , averaged over the whole spatial domain, i.e.

$$\text{SAM}(\mathcal{X}, \hat{\mathcal{X}}) = \frac{1}{WH} \sum_{j=1}^{WH} \arccos \frac{\hat{\mathbf{x}}_j^T \mathbf{x}_j}{\|\hat{\mathbf{x}}_j\|_2 \|\mathbf{x}_j\|_2}. \quad (56)$$

SAM is given in degrees. The smaller SAM, the less spectral distortions.

The third index is the degree of distortion (DD), which is defined as

$$\text{DD}(\mathcal{X}, \hat{\mathcal{X}}) = \frac{1}{WHS} \|\text{vec}(\mathcal{X}) - \text{vec}(\hat{\mathcal{X}})\|_1, \quad (57)$$

where $\text{vec}(\mathcal{X})$ and $\text{vec}(\hat{\mathcal{X}})$ denote the vectorizations of tensors \mathcal{X} and $\hat{\mathcal{X}}$, respectively, scaled to the range [0, 255]. DD is an indicator of the spectral quality of the estimated image. The smaller the DD, the better the spectral quality.

The fourth index is the relative dimensionless global error in synthesis (ERGAS), proposed in [56] and defined as

$$\text{ERGAS}(\mathcal{X}, \hat{\mathcal{X}}) = \frac{100}{c} \sqrt{\frac{1}{S} \sum_{i=1}^S \frac{\text{MSE}(\mathbf{X}^i, \hat{\mathbf{X}}^i)}{\mu_{\hat{\mathbf{X}}^i}^2}}, \quad (58)$$

where c is spatial downsampling factor, \mathbf{X}^i and $\hat{\mathbf{X}}^i$ denote the i^{th} band images of \mathcal{X} and $\hat{\mathcal{X}}$, respectively, $\text{MSE}(\mathbf{X}^i, \hat{\mathbf{X}}^i)$ is the

TABLE I
QUANTITATIVE RESULTS (RMSE, SAM, DD, ERGAS, UIQI, TIME) OF THE TEST METHODS ON THE PAVIA UNIVERSITY [54].

Methods	Pavia University [54]					
	RMSE	SAM	DD	ERGAS	UIQI	Time(s)
Best Values	0	0	0	0	1	0
GSOMP [24]	2.917	2.504	1.640	0.815	0.988	431
HySure [26]	1.953	2.059	1.237	0.515	0.993	43
BSR [19]	2.197	2.136	1.340	0.598	0.993	1437
NLSTF [30]	1.972	1.975	1.180	0.555	0.993	41
CSTF	1.747	1.782	1.055	0.485	0.995	64

TABLE II
QUANTITATIVE RESULTS (RMSE, SAM, DD, ERGAS, UIQI, TIME) OF THE TEST METHODS ON THE CUPRITE MINE [55].

Methods	Cuprite Mine [55]					
	RMSE	SAM	DD	ERGAS	UIQI	Time(s)
Best Values	0	0	0	0	1	0
GSOMP [24]	3.219	1.270	1.649	1.296	0.957	429
HySure [26]	2.919	1.177	1.634	1.170	0.965	42
BSR [19]	3.861	1.535	2.479	1.400	0.933	5051
NLSTF [30]	3.473	1.102	2.206	0.992	0.931	50
CSTF	2.507	0.991	1.280	0.975	0.966	99

mean square error between \mathbf{X}^i and $\hat{\mathbf{X}}^i$, and $\mu_{\hat{\mathbf{X}}^i}$ is the mean value of $\hat{\mathbf{X}}^i$. ERGAS reflects the overall quality of the fused image. The smaller ERGAS, the better the fusion results.

The fifth index is based on the Universal Image Quality Index (UIQI), proposed by Wang *et al.* [57]. It is computed on a sliding window of size 32×32 pixels and averaged over the spatial domain. Let \mathbf{X}_j^i and $\hat{\mathbf{X}}_j^i$ denote the j^{th} window of i^{th} band ground truth image and estimated image, respectively. The UIQI between i^{th} band images \mathbf{X}^i and $\hat{\mathbf{X}}^i$ is given by

$$\text{UIQI}(\mathbf{X}^i, \hat{\mathbf{X}}^i) = \frac{1}{M} \sum_{j=1}^M \frac{\sigma_{\mathbf{X}_j^i} \hat{\mathbf{X}}_j^i}{\sigma_{\mathbf{X}_j^i} \sigma_{\hat{\mathbf{X}}_j^i}} \frac{2\mu_{\mathbf{X}_j^i} \mu_{\hat{\mathbf{X}}_j^i}}{\mu_{\mathbf{X}_j^i} + \mu_{\hat{\mathbf{X}}_j^i}} \frac{2\sigma_{\mathbf{X}_j^i} \sigma_{\hat{\mathbf{X}}_j^i}}{\sigma_{\mathbf{X}_j^i} + \sigma_{\hat{\mathbf{X}}_j^i}}, \quad (59)$$

where M is the number of window positions, $\sigma_{\mathbf{X}_j^i} \hat{\mathbf{X}}_j^i$ is the sample covariance between \mathbf{X}_j^i and $\hat{\mathbf{X}}_j^i$, and $\mu_{\mathbf{X}_j^i}$ and $\sigma_{\mathbf{X}_j^i}$ are the mean value and standard deviation of \mathbf{X}_j^i , respectively. The definition of the UIQI index is extended to a HSI by averaging over all bands, i.e.

$$\text{UIQI}(\mathcal{X}, \hat{\mathcal{X}}) = \frac{1}{S} \sum_{i=1}^S \text{UIQI}(\hat{\mathbf{X}}^i, \mathbf{X}^i). \quad (60)$$

The UIQI combines the loss of correlation, luminance distortion, and contrast distortion. This index has a range of $[-1, 1]$, being equal to 1 when $\mathcal{X} = \hat{\mathcal{X}}$. The larger the value, the better the fusion results.

The sixth index is the computational time in seconds, which is used as the measure of computational efficiency. All the algorithms under test are implemented using Mathworks MATLAB R2016a, and the computer facility is equipped with Core-i5-6400 CPU with 2.7-GHz speed and 16-GB random access memory.

D. Parameters Discussions

To evaluate the CSTF sensitivity w.r.t. its key parameters, we run it for different values of the number iterations K , the weight of proximal term β , sparsity regularization parameter λ , and the number of atoms of dictionary of three modes n_w , n_h , n_s .

As summarized in **Algorithm 2**, we take the proximal alternative optimization scheme to solve the problem (15), and the estimation of all block is iterated. To evaluate the influence of the number of iterations K , we run CSTF for the different number of iterations K . Fig. 2 shows the RMSE of the reconstructed HSIs of Pavia University and Cuprite Mine as functions of K . As can be seen from the Fig. 2, the RMSE for Pavia University decreases when K varies from 1 to 6 and remains stable for $K > 6$. For Cuprite Mine, the RMSE has a sharp drop when K varies from 1 to 16, and then it decreases slowly. The maximum number of iterations is thus set as 20 for the CSTF.

Parameter β is the weight of proximal term in (16). Fig. 3 shows the RMSE as a function of $\log(\beta)$ for the Pavia University and Cuprite Mine. The values in the x-axis are given by $\log(\beta)$ (\log is base 10). As can be seen from Fig. 3, the RMSE for Pavia University keeps relatively stable, when $\log(\beta)$ varies from -5 to -3 , and then it rises when $\log(\beta)$ varies from -3 to 1 . The RMSE for Cuprite Mine keeps relatively stable when $\log(\beta)$ varies from -5 to -2 , and then it drops to minimum value as $\log(\beta) = -1$. Therefore, we set $\log\beta = -3$ and $\log\beta = -1$, that is, $\beta = 0.001$ and $\beta = 0.1$, for Pavia University and Cuprite Mine, respectively.

Parameter λ controls the sparsity of the core tensor \mathcal{C} and, therefore, affects the estimation of HR-HSI. Higher values of λ yields sparser core tensors \mathcal{C} . Fig. 4 shows the RMSE as functions of $\log(\lambda)$. The values in the x-axis are given by $\log(\lambda)$ (\log is base 10). As can be seen from Fig. 4, the RMSE for Pavia University and Cuprite Mine have a slight decrease as $\log(\lambda)$ varies from -6 to -5 , keeps constant for $\log(\lambda) \in [-5, -4]$, and increases sharply for $\log(\lambda) > -4$. Therefore, we set $\lambda = 10^{-5}$ for both Pavia University and Cuprite Mine.

Fig. 5 (a), Fig. 5 (b), and Fig. 5 (c) show the RMSE of the reconstructed Pavia University and Cuprite Mine as functions of the number of dictionary atoms n_w , n_h and n_s , respectively. As can be seen from Fig. 5 (a) and (b), the RMSE for Pavia University and Cuprite Mine has a sharp drop when n_w and n_h vary from 40 to 240. However, when n_w and n_h grow higher, the RMSE does not change obviously. Therefore, we set $n_w = 240$, $n_h = 240$ in the Pavia University and Cuprite Mine. As Fig. 5 (c) shows, the RMSE for Pavia University decreases as n_s varies from 3 to 12, and then it will rise as n_s increases further. For Cuprite Mine, the RMSE curve decreases as n_s varies from 3 to 12, and then it does not change obviously as n_s increases further. Therefore, we set $n_s = 12$ for both Pavia University and Cuprite Mine. The CSTF for Pavia University and Cuprite Mine needs larger dictionaries of width and height modes, and smaller dictionaries of spectral mode. The reason is that the spectral signatures of both HSIs live on low dimensional subspaces [40].

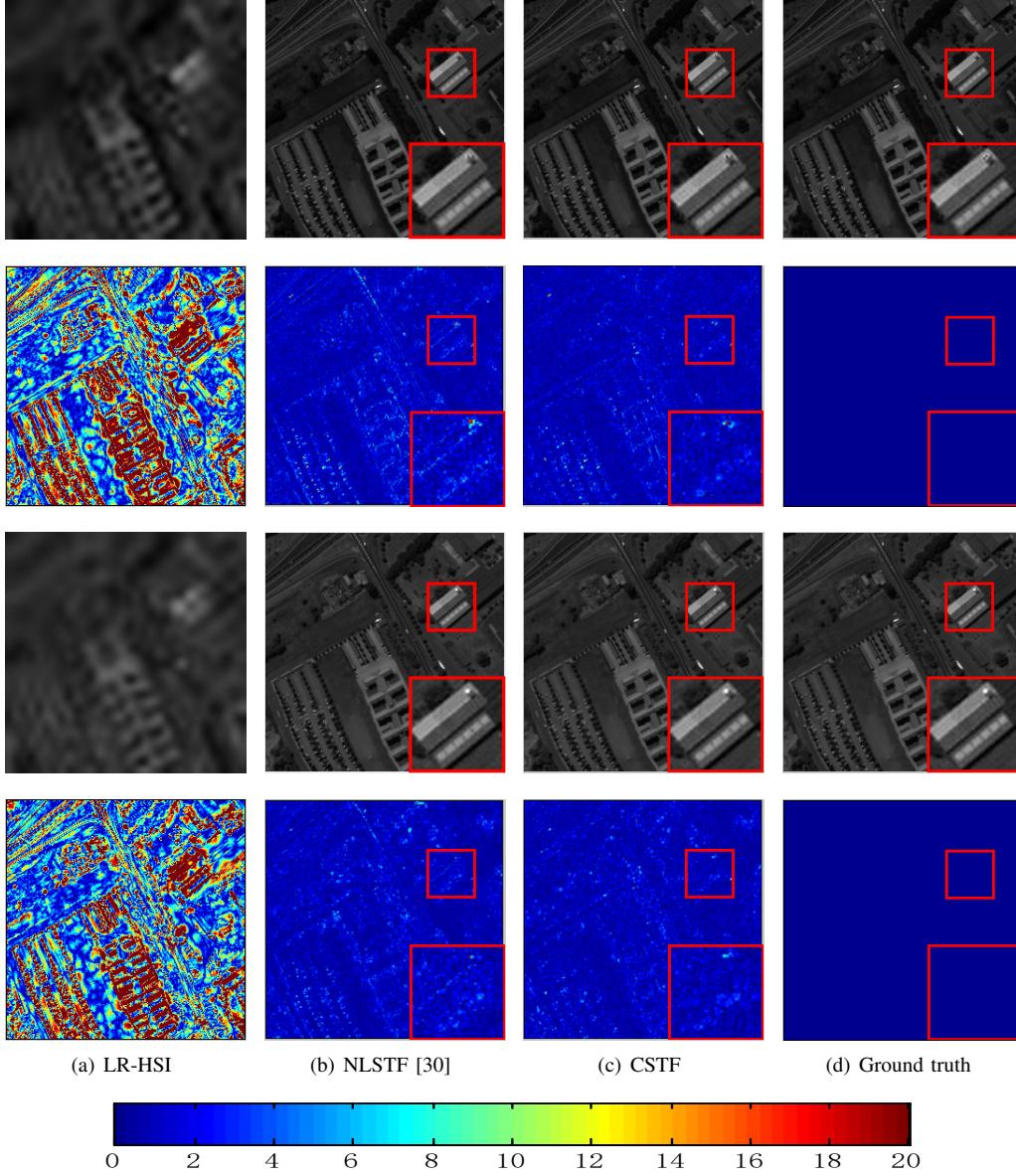


Fig. 6. Reconstructed images and corresponding error images of Pavia University for the 40th and 60th bands with a downsampling factor 8. The first and second rows show the reconstructed images for the 40th band corresponding error images, respectively; The third and forth rows show the reconstructed images for the 60th band and corresponding error images, respectively; (a) LR-HSI; (b) the NLSTF method [30]; (c) the proposed CSTF method ; (d) Ground truth.

E. Experimental Results

In this section, we show the fusion result of the five test methods on the Pavia University and Cuprite Mine.

Table I shows the RMSE, SAM, DD, ERGAS, UIQI, and running time of the recovered HSIs for the Pavia University. The best results are marked in bold for clarity. As can be seen from Table I, CSTF method performs better among the compared methods in terms of reconstruction accuracy, and NLSTF and HySure are the fastest methods. CSTF and NLSTF methods are very effective in reconstructing the Pavia University, with the CSTF method outperforming NLSTF. In order to further compare the performance of the proposed CSTF method, the reconstructed 40th and 60th bands and corresponding error images yielded by NLSTF and CSTF at

the are shown in Fig. 6. The error images reflect the differences between the fusion results and ground truths. For visual comparison, a meaningful region for each of the resulting image is magnified. As can be seen from the magnified image region, the NLSTF and CSTF methods provide clear and sharp spatial details compared with the LR-HSI. NLSTF produces a few spectral distortions at 40th band, and CSTF performs slightly better than the NLSTF.

Table II shows the average objective results for the Cuprite Mine in terms of RMSE, SAM, DD, ERGAS, and UIQI. As it can be seen from Table II, CSTF and HySure perform consistently better than the other compared methods. The NLSTF performance is relatively low in this dataset.

TABLE III
QUANTITATIVE RESULTS OF NOISY CASES ON THE PAVIA UNIVERSITY [54].

Methods	SNRh=35dB, SNRm=40dB					SNRh=30dB, SNRm=35dB				
	RMSE	SAM	DD	ERGAS	UIQI	RMSE	SAM	DD	ERGAS	UIQI
Best Values	0	0	0	0	1	0	0	0	0	1
GSOMP [24]	2.985	2.701	1.803	0.839	0.987	3.227	3.073	2.101	0.921	0.983
HySure [26]	2.058	2.112	1.319	0.561	0.992	2.294	2.455	1.586	0.632	0.991
BSR [19]	2.195	2.254	1.405	0.612	0.992	2.392	2.448	1.613	0.675	0.991
NLSTF [30]	2.223	2.404	1.498	0.627	0.991	2.625	2.694	1.790	0.810	0.988
CSTF	1.961	2.048	1.290	0.550	0.993	2.152	2.271	1.479	0.608	0.991

F. Experimental Results of Noisy Case

In practice, there is additive noise in the hyperspectral and multispectral imaging process, and the noise level in LR-HSI is often higher than that of HR-MSI. To test the robustness of the proposed CSTF method to the noise, we firstly simulate the \mathcal{Y} and \mathcal{Z} in the same way as the previous experiments for the Pavia University and then add Gaussian noise to the LR-HSI and HR-MSI. The noisy LR-HSI and HR-MSI are fused to produce the HR-HSI. The SNR of noisy LR-HSI and HR-MSI are denoted as SNRh and SNRm, respectively. Table III shows the quality metric values of the noisy cases on the Pavia University. We can see that the CSTF method still outperforms the other testing methods in the noisy case.

VI. CONCLUSIONS

In this paper, we present a novel coupled sparse tensor factorization based framework, termed as CSTF, to estimate an HR-HSI, by fusing an LR-HSI with an HR-MSI counterpart. Unlike recent matrix factorization based HSI-MSI fusion methods, the proposed CSTF method considers the HSI as a tensor with three modes and factorizes it as a sparse core tensor multiplication by dictionaries of three modes. The proposed CSTF is compared with the state-of-the-art methods on two semi-real remote sensing HSIs. The obtained results systematically outperform the competitors, providing experimental evidence of the effectiveness of the proposed CSTF method.

In our future works, we aim to extend the method in two directions. On one hand, the intensities of the HSI are naturally non-negative. Thus, by adding non-negative constraints to the tensor factorization, the performance of the method is expected to be further improved. On the other hand, there are non-local spatial similarities in the HSI, which has been proved to be effective for many image restoration problems. Thus, the incorporation of the non-local spatial similarities into the CSTF method will be investigated.

ACKNOWLEDGEMENT

The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions, which have significantly improved this paper. They would also like to thank Dr. N. Akhtar for providing the codes of BSR and GSOMP.

REFERENCES

- [1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [2] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, to be published, 2018, DOI:10.1109/TGRS.2018.2794326.
- [3] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.
- [4] L. Fang, N. He, S. Li, P. Ghamsari, and J. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
- [5] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiotemporal information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.
- [6] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.
- [7] G. Shaw and H. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, Jan. 2003.
- [8] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, 2018, DOI:10.1109/TNNLS.2018.2798162.
- [9] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Bruce, "Comparison of pan-sharpening algorithms: Outcome of the 2006 GRSS data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [10] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.
- [11] C. Chen, Y. Li, W. Liu, and J. Huang, "SIRF: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4213–4224, Nov. 2015.
- [12] F. Fang, F. Li, C. Shen, and G. Zhang, "A variational approach for pansharpening," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2822–2834, Jul. 2013.
- [13] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.
- [14] R. Restaino, G. Vivone, M. Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.
- [15] Z. Chen, H. Pu, B. Wang, and G. M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1418–1422, Aug. 2014.
- [16] L. Loncan, L. Almeida, J. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simões, J. Tourneret, M. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [17] R. Hardie, M. Eismann, and G. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.
- [18] Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

- [19] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 3631–3640.
- [20] R. Kawakami, J. Wright, Y. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 2329–2336.
- [21] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled non-negative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [22] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3586–3594.
- [23] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J. Tourneret, M. Chen, and S. Godsill, "Multi-band image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.
- [24] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Euro. Conf. Comput. Vis.*, Sep. 2014, pp. 63–78.
- [25] M. Viganjones, M. Simões, G. Licciardi, N. Yokoya, J. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 274–288, Jan. 2016.
- [26] M. Simões, J. Bioucas-Dias, L. 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, Jun. 2015.
- [27] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [28] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
- [29] M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [30] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *IEEE Conf. Comput. Vis. Pattern Recog.*, Jul. 2017, pp. 3862–3871.
- [31] J. Bengua, H. Phien, H. Tuan, and M. Do, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2466–2479, May 2017.
- [32] T. Jiang, T. Huang, X. Zhao, T. Ji, and L. Deng, "Matrix factorization for low-rank tensor completion using framelet prior," *Inf. Sci.*, vol. 436, pp. 403–417, 2018.
- [33] T. Jiang, T. Huang, X. Zhao, L. Deng, and Y. Wang, "A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors," in *IEEE Conf. Comput. Vis. Pattern Recog.*, Jul. 2017, pp. 4057–4066.
- [34] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1030–1040, Apr. 2018.
- [35] C. Jia and Y. Fu, "Low-rank tensor subspace learning for RGB-D action recognition," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4641–4652, Oct. 2016.
- [36] C. Jia, M. Shao, and Y. Fu, "Sparse canonical temporal alignment with deep tensor decomposition for action recognition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 738–750, Feb. 2017.
- [37] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017.
- [38] L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 23, no. 8, pp. 3336–3351, Sep. 1996.
- [39] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [40] L. Zhuang and J. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
- [41] Y. Rivenson and A. Stern, "Compressed imaging with a separable sensing operator," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 449–452, Jun. 2009.
- [42] T. Lu, S. Li, L. Fang, Y. Ma, and J. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.
- [43] R. Arablouei and F. Hoog, "Hyperspectral image recovery via hybrid regularization," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5649–5663, Dec. 2016.
- [44] L. Zhang, W. Wei, C. Tian, F. Li, and Y. Zhang, "Exploring structured sparsity by a reweighted laplace prior for hyperspectral compressive sensing," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4974–4988, Oct. 2016.
- [45] L. Fang, C. Wang, S. Li, and J. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.
- [46] H. Attouch, J. Bolte, and B. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Math. Program.*, vol. 137, no. 1-2, pp. 91–129, Feb. 2013.
- [47] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, Apr. 2010.
- [48] G. Golub and C. Van, *Matrix computations*. JHU Press, 2012.
- [49] O. Axelsson, *Iterative solution methods*. Cambridge University Press, 1996.
- [50] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.
- [51] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, Jul. 2011.
- [52] J. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Hyperspectral Image and Signal Process. Workshop: Evolution in Remote Sens.*, Aug. 2009, pp. 1–4.
- [53] L. Smith and M. Elad, "Improving dictionary learning: Multiple dictionary updates and coefficient reuse," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 79–82, Jan. 2013.
- [54] F. Dell'Acqua, P. Gamba, A. Ferrari, J. Palmason, and J. Benediktsson, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.
- [55] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, and M. Solis, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
- [56] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Jan. 2000, pp. 99–103.
- [57] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.