NSTMR: Super Resolution of Sentinel-2 Images Using Nonlocal Nonconvex Surrogate of Tensor Multirank

Xuan-Qi Wang and Teng-Yu Ji

***Abstract*—In this article, we address the super-resolution problems, which estimate the high-resolution multispectral images from the multispectral Sentinel-2 (S2) images with different resolutions. Since S2 images can be naturally represented by tensors, we reformulate the degradation process as the tensor-based form. Based on the degradation mechanism, we build a tensor-based optimization model for S2 images super-resolution problem, which fully exploits intrinsic nonlocal spatial similarity and global spectral redundancy. Specifically, the model consists of the data fidelity term and the low-multirank regularizer tailored to thoroughly mining the inherent spatial-nonlocal and spectral redundancy. Then, we develop an efficient alternating direction method of multipliers algorithm with theoretically guaranteed convergence to tackle the resulting tensor optimization problem. Numerical experiments including simulated and real data demonstrate that our method outperforms the competing methods visually and qualitatively.**

***Index Terms*—Alternating direction method of multipliers (ADMMs), global spectral redundancy, nonlocal spatial similarity, sentinel-2 (S2) image.**

I. Introduction

IN REMOTE sensing, an increasing number of satellites are launched to acquire multispectral (MS) images, which are used to perform terrestrial observations in support of services such as environmental monitoring, land cover changes detection, and natural disaster management [1]–[3]. However, due to the restrictions of imaging gadgets, there is a tradeoff among spatial and spectral resolutions of MS images, i.e., the spatial resolution [or ground sampling distance (GSD)] of images acquired by sensors varies according to different spectral bands. To obtain a higher signal-to-noise ratio (SNR), the spatial resolution has to be lower if the spectral one is required to be higher, so optical images might be blurry. In contrast, losing spectral resolution is the price to pay for a high spatial resolution. For tradeoff, many satellite sensors, such as MODIS, ASTER, and Sentinel-2 (S2), have been invented to get low-spatial-resolution MS bands together with some high-spatial-resolution bands (panchromatic image). Therefore, it is desirable to generate spatially enhanced MS data by exploiting the structure in the panchromatic image. As a fundamental task in remote sensing, super resolution (SR) for multispectral and multiresolution images aims to infer the data of all the bands given different resolutions. Without loss of generality, in this article, we focus on the SR on S2 dataset. S2 is a multispectral operational imaging mission operated by the European Space Agency (ESA) [4], [5]. It acquires MS images composed of 13 bands (443–2190 nm) in the visible-near infrared (VNIR), and shortwave-infrared (SWIR) spectrum at three different spatial resolutions; see Table I for more details. There are four bands at 10 m (VNIR), six at 20 m (SWIR), and three at 60 m. Band B10 is discarded since this band is to detect Cirrus clouds and does not provide any structures of the ground surfaces, thus, we do not process this band.

This article aims to infer all bands of the S2 images at 20 m and 60 m spatial resolutions, such that all bands have the same and maximal resolution (10 m). Fig. 1 demonstrates the SR process on synthetic S2 images and the results of our algorithm.

Existing SR on S2 images can be roughly categorized into three classes: pansharpening methods [6]–[11], deep learning methods [12]–[14], and model-based methods [15]–[19]. Pansharpening methods [6]–[11] fuse a high-spatial-resolution channel (i.e., a panchromatic image) with other low-spatial-resolution MS bands. This strategy has also been utilized to S2, though the sensors of S2 do not fully satisfy the demand since its sensors do not have a panchromatic band that covers most spectral ranges and they only have four channels at 10 m bands (VNIR). The pansharpening methods can be used to address the S2 problem by creating a single high-resolution (HR) band using 10 m bands and then sharpening the 20 m and 60 m bands. Du *et al.* [9] compared pansharpening algorithms, including principal component analysis [20], high pass filter, intensity hue saturation [21], and À-trous wavelet transforms for upscaling the 20 m SWIR bands to 10 m spatial resolution. Vaiopoulos and Karantzalos [10] comprehensively evaluated the performance of 21 pansharpening algorithms on enhancing 20 m SWIR and VNIR bands to 10 m spatial resolution. Wang *et al.* [11] presented the area-to-point regression kriging (ATPRK) method to sharpen the 20 m bands to 10 m spatial resolution. For these

Manuscript received January 29, 2021; revised April 15, 2021; accepted May 18, 2021. Date of publication May 25, 2021; date of current version June 10, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 12001432, and in part by the Fundamental Research Funds for the Central Universities under Grant 31020180QD126. *(Corresponding author: Teng-Yu Ji.)*

Xuan-Qi Wang is with the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710072, China, and also with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: xuanqi_wang@outlook.com).

Teng-Yu Ji is with the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: tengyu_j66@126.com).

Digital Object Identifier 10.1109/JSTARS.2021.3083495

TABLE I
13 SENTINEL-2 BANDS

Band	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Center wavelength (nm)	443	490	560	665	705	740	783	842	865	945	1380	1610	2190
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

TABLE II
TENSOR NOTATIONS

Notations	Explanations
$a, \mathbf{a}, \mathbf{A}, \boldsymbol{\mathcal{A}}$	Scalar, vector, matrix, tensor.
$\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}}$	The **t-product** of two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$.
$\|\boldsymbol{\mathcal{A}}\|_F$	The **Frobenius norm** of a tensor $\boldsymbol{\mathcal{A}}$.
$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle$	The **inner product** of two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$.
$\|\boldsymbol{\mathcal{A}}\|_{\text{TNN}}$	The **Tensor nuclear norm** of a tensor $\boldsymbol{\mathcal{A}}$.
$\boldsymbol{\mathcal{A}} \odot \boldsymbol{\mathcal{B}}$	Elementwise multiplication (Hadamard product).
$\boldsymbol{\mathcal{A}}./\boldsymbol{\mathcal{B}}$	Elementwise division.

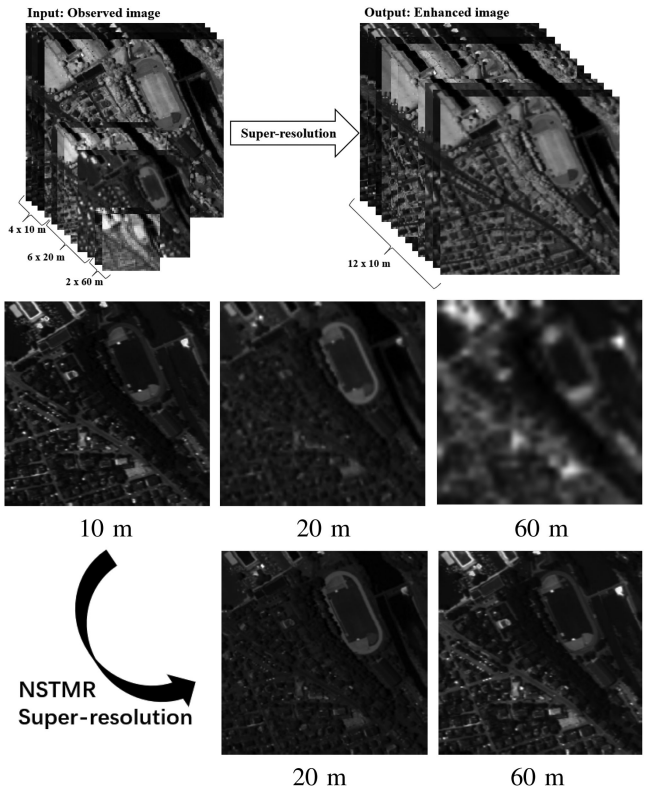

Fig. 1. Input and output of super resolution on APEX images. Top: Input and output data by NSTMR in 3-D vision (the bands are reordered for better visualization). Middle: Input bands at 10 m, 20 m, and 60 m GSD. Bottom: Super-resolved bands to 10 m GSD by the proposed method (NSTMR).

methods, the main weakness is that much information in HR bands cannot be fully employed in the fusion.

The recent deep learning methods try to learn a mapping between lower resolution input and higher resolution output from texture and modality of example data. Some neural networks [12]–[14] take low-resolution (LR) images as input to recover all HR images. Such learning-based methods highly depend on the diversity and credibility of training datasets.

Model-based methods [15]–[19] formulate the super-resolution problem as an ill-posed inverse problem and then estimate the HR images under the variational regularization framework by solving an optimization problem that super resolves all bands simultaneously. These methods rely on an observation model, which generally contains two terms: the fidelity term and the regularization term. The former derives from the process of blurring, downsampling, and adding noise and the latter describes image prior to mitigate the ill-posedness of the inverse problems. Brodu [15] first upscaled 20 m and 60 m bands to 10 m spatial resolution by preserving band-dependent information (reflectance) and propagating band-independent information to preserve the subpixel details. By exploiting the spectral redundancy of S2 images, Lanaras *et al.* [16] proposed a low subspace-based model with total variation regularization for jointly inferring all bands at 10 m spatial resolution. Paris *et al.* [17] presented a similar approach, which incorporates the block match three-dimensional (3-D) (BM3D) denoiser. Ulfarsson *et al.* [18] relied on the projections onto a low-dimensional subspace that is automatically estimated during the optimization procedure and got the whole images of all bands. Lin and Bioucas-Dias [19] adopted the same framework but introduced another prior knowledge, i.e., the self-similarity graph learned from the original images. However, the key issue of these methods is to exploit the prior of high-dimensional data in the matrix or vector forms as opposed to tensor form. That means these methods cannot grasp intrinsic structures of high-dimensional images, such as global low-rankness, the features of local and nonlocal self-similarity. Inspired by this, we propose a novel method, which takes a nonlocal low tensor multirank prior into account.

Recently, most model-based methods employ the nonlocal self-similarity priors for image processing problems [22]–[25]. However, these methods cannot preserve the intrinsic properties of these images in the matrix form. Meanwhile, tensor analysis has played an increasingly important role in the SR process from LR images to HR images [26]–[31]. LR images can be easily represented by a 3-D tensor and it is natural to super-resolve LR images from the tensor perspective. Some tensor factorization for 3-D images, including the tensor singular value decomposition (t-SVD), are used in super resolution reconstruction problems [32]–[34]. For S2 data, they can be naturally represented by tensors, so we reformulate the degradation process as the tensor-based form. The prior information in the matrix form cannot describe the spatial correlation of S2 image. However, since different bands have redundant features in the spatial and spectral dimension and these features can be characterized by low-rankness [30, 35, 36]. The use of tensor can better retain the redundancy and correlation of spatial and spectral dimensions. In

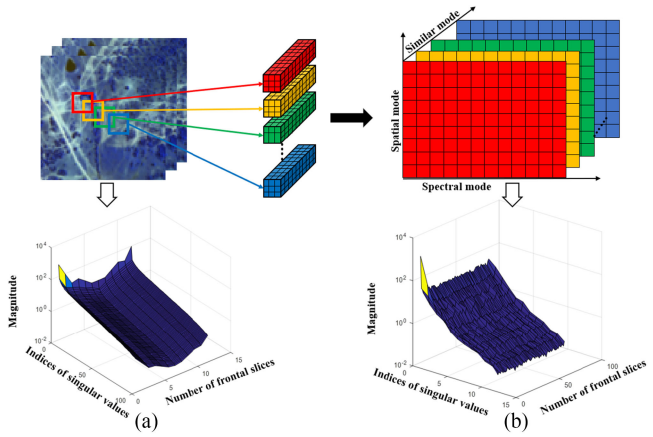

Fig. 2. Top: The flow diagram of the construction of the tensor by stacking similar 3-D patches. Bottom: the singular values of original tensor and the tensor stacked by similar 3-D patches after the Fourier transform along the third mode.

other words, the low-rankness property can be better depicted by tensor-based model. Prvost *et al.* [35] took advantage of coupled Tucker approximation.

Moreover, tensors stacked by nonlocal similar 3-D patches can characterize the low-rankness more clearly, in particular, t-SVD-based multirank can retain the intrinsic structure of S2 images. For an original MS image, its singular value distribution is shown in the Fig. 2(a) and the distribution after finding the nonlocal similar 3-D patches and aggregating them is shown in Fig. 2(b). We can see that the magnitude of the singular values in Fig. 2(b) attenuate more quickly, showing the low-rankness is characterized more obviously. In addition, tensor nuclear norm (TNN) gives convex relaxation of multirank but it is not an accurate approximation since it treats each singular value equally, causing the suboptimal solutions. To address this issue, we use an improved low-rank sparsity measure to approximate the tensor multirank more accurately, such as the nonconvex logdet surrogate. This method has smaller shrinkage for larger singular values and larger shrinkage for smaller singular values and is easy to calculate.

According to these two facts, in this article, we propose a method using nonlocal nonconvex surrogate of tensor multirank (NSTMR) for S2 image super resolution. The contributions of our work are mainly three folds as follows.

1) We suggest a S2 SR model whose data fidelity term depicts the tensor-based degradation process and the regularization term fully exploits the nonlocal spatial similarity and global spectral redundancy of S2 images by using a logdet-based nonconvex surrogate of tensor multirank regularizer.
2) We develop an alternating direction method of multipliers (ADMM) to solve the proposed model efficiently and establish the theoretical guarantee of its global convergence to a sandle point of the argument Largrangian function.
3) Experiments on the simulated and real data are conducted to demonstrate the superior performance of the proposed algorithm in contrast with state-of-the-art alternatives.

The outline of this article is as follows. Section II gives some preliminary knowledge and notations on tensor analysis. Then, Section III presents the formulation of our model together with the solving algorithm. Section IV covers experimental results and Section V discusses the convergence and parameter analysis of the NSTMR. Finally, Section VI concludes this article.

II. Notations and Preliminaries

A. Notations

Scalars are denoted by lowercase letters, e.g., a, vectors by boldface lowercase letters, e.g., $\mathbf{a}$, matrices by boldface uppercase letters, e.g., $\mathbf{A}$, tensors by Euler script letters, e.g., $\boldsymbol{\mathcal{A}}$. Given tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, its entries are denoted by $a_{i,j,k}$ for $1 \le i \le I, 1 \le j \le J, 1 \le k \le K$. For $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I\times J\times K}$, the inner product $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle$ is $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i,j,k} a_{i,j,k} b_{i,j,k}$, the Hadamard product (elementwise multiplication) $\boldsymbol{\mathcal{A}} \odot \boldsymbol{\mathcal{B}}$ is $(\boldsymbol{\mathcal{A}} \odot \boldsymbol{\mathcal{B}})_{i,j,k} = a_{i,j,k} b_{i,j,k}$, similarly, $(\boldsymbol{\mathcal{A}}./\boldsymbol{\mathcal{B}})_{i,j,k} = a_{i,j,k}/b_{i,j,k}$ is the elementwise division of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. The Frobenius norm of $\boldsymbol{\mathcal{A}}$ is defined as $\|\boldsymbol{\mathcal{A}}\|_F := \sqrt{\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}} \rangle}$. The $\mathtt{vec}(\mathbf{A})$ is used to vectorize the matrix $\mathbf{A}$. For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, the ith frontal slice is denoted as $\boldsymbol{\mathcal{A}}^{(i)}$, and $\overline{\boldsymbol{\mathcal{A}}}$ is the result of discrete Fourier transformation of along the third-dimension by using the MATLAB command `fft`, i.e., $\overline{\boldsymbol{\mathcal{A}}}$ = `fft`($\boldsymbol{\mathcal{A}}$,[],3). In the same way, $\boldsymbol{\mathcal{A}}$ can be obtained from $\overline{\boldsymbol{\mathcal{A}}}$ via `ifft`($\overline{\boldsymbol{\mathcal{A}}}$,[],3) that denotes the inverse Fourier transformation. The conjugate transpose of $\boldsymbol{\mathcal{A}}$ is defined as $\boldsymbol{\mathcal{A}}^T \in \mathbb{R}^{J\times I\times K}$ in [37], and the identity tensor $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{I\times I\times K}$ is the tensor whose first frontal slice is identity matrix and other frontal slices are zero. The f-diagonal tensor satisfies every frontal slice is a diagonal matrix. An orthogonal tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times I\times K}$ satisfies $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{A}}^T * \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{I}}$. $\mathbf{1}_{I\times J\times K}$ and $\mathbf{0}_{I\times J\times K}$ denotes the $I \times J \times K$ tensors whose entries are 1 and 0, respectively, and the subscript is omitted without causing confusion.

Definition 2.1 (t-product [37]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{J\times L\times K}$, the t-product $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}}$ is the $I \times L \times K$ tensor

$$\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} = \mathtt{Fold}(\mathtt{bcirc}(\boldsymbol{\mathcal{A}})\mathtt{Unfold}(\boldsymbol{\mathcal{B}}))$$

where

$$\mathtt{bcirc}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{\mathcal{A}}^{(1)} & \boldsymbol{\mathcal{A}}^{(K)} & \boldsymbol{\mathcal{A}}^{(K-1)} & \dots & \boldsymbol{\mathcal{A}}^{(2)} \\ \boldsymbol{\mathcal{A}}^{(2)} & \boldsymbol{\mathcal{A}}^{(1)} & \boldsymbol{\mathcal{A}}^{(K)} & \dots & \boldsymbol{\mathcal{A}}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{A}}^{(K)} & \boldsymbol{\mathcal{A}}^{(K-1)} & \boldsymbol{\mathcal{A}}^{(K-2)} & \dots & \boldsymbol{\mathcal{A}}^{(1)} \end{bmatrix}$$

and $\mathtt{Unfold}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{\mathcal{A}}^{(1)} \\ \boldsymbol{\mathcal{A}}^{(2)} \\ \vdots \\ \boldsymbol{\mathcal{A}}^{(m_3)} \end{bmatrix}$, $\mathtt{Fold}(\mathtt{Unfold}(\boldsymbol{\mathcal{A}})) = \boldsymbol{\mathcal{A}}$.

Definition 2.2 (t-SVD [37]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, the t-SVD decomposition of the tensor $\boldsymbol{\mathcal{A}}$ is

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{I\times I\times K}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{J\times J\times K}$ are orthogonal, and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{I\times J\times K}$ is an f-diagonal tensor.

Definition 2.3 (Tensor multirank [38]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, the tensor multirank is a vector $\mathbf{v} \in \mathbb{R}^K$, whose jth element is the rank of jth frontal slice of $\overline{\boldsymbol{\mathcal{A}}}$, i.e., $\mathbf{v}_j = \mathtt{rank}(\overline{\boldsymbol{\mathcal{A}}}^{(j)})$.

Definition 2.4 (Tensor nuclear norm [38]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I\times J\times K}$, the sum of the singular values of all frontal slices of $\overline{\boldsymbol{\mathcal{A}}}$ is defined as the tensor nuclear norm $||\boldsymbol{\mathcal{A}}||_{\mathbf{TNN}}$, i.e.,

$$\|\boldsymbol{\mathcal{A}}\|_{\mathbf{TNN}} = \sum_{i=1}^{m_3}\sum_{j}\sigma_j\left(\overline{\boldsymbol{\mathcal{A}}}^{(i)}\right)$$

where $\sigma_j(\overline{\boldsymbol{\mathcal{A}}}^{(i)})$ is the jth singular value of $\overline{\boldsymbol{\mathcal{A}}}^{(i)}$. Compared with the tensor nuclear norm, which treats each singular value equally and gets the sum of singular values, the log-sum of singular values has more enhanced performance [39]. To approximate tensor multirank accurately, we consider logdet function-based nonconvex surrogate of tensor multirank as

$$\|\boldsymbol{\mathcal{A}}\|_{\mathrm{NSTMR}} = \sum_{i=1}^{m_3}\sum_{j}\log\left(\sigma_j(\overline{\boldsymbol{\mathcal{A}}}^{(i)}) + \varepsilon\right)$$

where ε is a very small positive number.

Definition 2.5 (Twist and squeeze operation [38]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{M\times 1\times N}$, it can be obtained using the twist operation acting on the matrix $\mathbf{B}$ of size $M \times N$. In the meanwhile, a $M \times N$ matrix $\mathbf{B}$ can be obtained using the squeeze operation on $\boldsymbol{\mathcal{A}}$, i.e.,

$$\boldsymbol{\mathcal{A}} = \mathtt{twist}(\mathbf{B}), \mathbf{B} = \mathtt{squeeze}(\boldsymbol{\mathcal{A}}).$$

For an MS image $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{M\times N\times B}$, it can be obtained by twist, convolution, and squeeze on $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M\times N\times B}$ successively for each band image, i.e.,

$$[\boldsymbol{\mathcal{C}}^{(1)}, \ldots, \boldsymbol{\mathcal{C}}^{(b)}] = [\mathtt{squeeze}(\boldsymbol{\mathcal{B}}_1 * \mathtt{twist}(\boldsymbol{\mathcal{X}}^{(1)})), \ldots, \mathtt{squeeze}(\boldsymbol{\mathcal{B}}_b * \mathtt{twist}(\boldsymbol{\mathcal{X}}^{(b)})] \quad (1)$$

where $\boldsymbol{\mathcal{B}}_i \in \mathbb{R}^{M\times M\times N}, i \in \{1, \ldots, b\}$ is the block circulant with circulant blocks blurring operator under the periodic boundary conditions.

Definition 2.6: We define the MS blurring procedure as follows:

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}} \quad (2)$$

where $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{M\times MB\times N}$ is the third order tensor whose column block tensors are composed of $\boldsymbol{\mathcal{B}}_i$, i.e., $\boldsymbol{\mathcal{B}} = [\boldsymbol{\mathcal{B}}_1, \ldots, \boldsymbol{\mathcal{B}}_b]$, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M\times N\times B}$ is the original MS image, and $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{M\times N\times B}$ is the blurred MS image which is the same with (1), i.e., $\boldsymbol{\mathcal{C}}^{(i)} = \mathtt{squeeze}(\boldsymbol{\mathcal{B}}_i * \mathtt{twist}(\boldsymbol{\mathcal{X}}^{(i)})$ for ith band.

The following gives an account of the selection and application of nonlocal similarity patches. We use the following strategy to construct $\boldsymbol{\mathcal{G}}_i$: First, we divide the spatial domain of the upsampled MSI of size $M \times N \times B$ into a set of full-band patches of size $\sqrt{W} \times \sqrt{W} \times B$ (space × space×bands) with overlapping sliding window strategy. Next, these patches are divided into n_t groups. For the ith group, it contains n_s patches that are the similar patches of ith reference patch of size $\sqrt{W} \times \sqrt{W} \times B$. The similar patches of the ith reference patch can be found in its neighboring area. Finally, we reshape each patch into matrix and stack the n_s matrices as tensor $\boldsymbol{\mathcal{G}}_i$ of size $W \times B \times n_s$ (see Fig. 2). The process of extracting $\boldsymbol{\mathcal{G}}_i \in \mathbb{R}^{W\times B\times n_s}$ from $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M\times N\times B}$ can be expressed as a mapping R_i, i.e., $\boldsymbol{\mathcal{G}}_i = R_i(\boldsymbol{\mathcal{X}})$. The operator $R_i^T : \mathbb{R}^{W\times B\times n_s} \to \mathbb{R}^{M\times N\times B}$ denotes the adjoint operator of R_i, i.e., $\langle R_i(\boldsymbol{\mathcal{X}}), \boldsymbol{\mathcal{G}}_i\rangle = \langle \boldsymbol{\mathcal{X}}, R_i^T(\boldsymbol{\mathcal{G}}_i)\rangle$ for any $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M\times N\times B}$. The adjoint operator $R_i^T(\boldsymbol{\mathcal{G}}_i)$ puts back the elements of $\boldsymbol{\mathcal{G}}_i$ into the tensor $\mathbf{0}$ of size $M \times N \times B$, where the positions of the elements of $\boldsymbol{\mathcal{G}}_i$ in $\mathbf{0}$ is consistent with the positions in $\boldsymbol{\mathcal{X}}$. The composition of R_i^T and R_i yields $R_i^T(R_i(\boldsymbol{\mathcal{X}})) = \boldsymbol{\mathcal{W}}_i \odot \boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{W}}_i$ of size $M \times N \times B$ computes the number of times that each pixel of $\boldsymbol{\mathcal{X}}$ occurs in $\boldsymbol{\mathcal{G}}_i$, i.e., $\boldsymbol{\mathcal{W}}_i = R_i^T(\mathbf{1}_{W\times B\times n_s})$, where $\mathbf{1}_{W\times B\times n_s}$ denotes the tensor with all ones of $W \times B \times n_s$.

III. Model and Algorithm

A. Problem Formulation

The degradation process can be written as the tensor form

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}}) + \boldsymbol{\mathcal{N}} \quad (3)$$

where $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M\times N\times B}$ is the MS ground truth image, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{M\times N\times B}$ is the MS image with the same spatial resolution after upsampling the jagged S2 observation images, $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{M\times N\times B}$ is the binary mask tensor, $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{M\times MB\times N}$ is the blurring tensor, and $\boldsymbol{\mathcal{N}}$ is the additive Gaussian noise. The downsampling is uniform and has a factor $d_i = 1, 2$, and 6 for 10 m, 20 m, and 60 m resolution, respectively. Altough $\boldsymbol{\mathcal{D}}$ represents the downsampling process, it has the same dimension with $\boldsymbol{\mathcal{X}}$ so that $\boldsymbol{\mathcal{Y}}$ is the same size with $\boldsymbol{\mathcal{X}}$ since $\boldsymbol{\mathcal{Y}}$ is the unsampled S2 images.

B. Proposed Model and Algorithm

To estimate the sharpened images $\boldsymbol{\mathcal{X}}$, we solve the optimization problem

$$\min_{\boldsymbol{\mathcal{X}}} \tfrac{1}{2}\|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}})\|_F^2 + \lambda\sum_{i=1}^{n_t}\|R_i(\boldsymbol{\mathcal{X}})\|_{\mathrm{NSTMR}} \quad (4)$$

where λ is the nonnegative regularization parameter, $R_i(\boldsymbol{\mathcal{X}})$ is the tensor formed by stacking similar patches for the ith reference patch, and n_t is the total number of groups of similar patches. First, the data-fitting term accounts for the individual blur and downsampling per band. Second, the nonconvex surrogate of tensor multirank term is used to exploit the spatial, spectral, and nonlocal redundancy of multispectral images. The proposed model (4) is denoted as "NSTMR" and the tensor form is used because it can ensure that the similar 3-D patches found preserve intrinsic characteristics.

To solve the abovementioned problem,we use ADMM algorithm [40]–[42]. By introducing auxiliary variables $\boldsymbol{\mathcal{P}}$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}$, we reformulate the optimization (4) as the equivalent constrained version

$$\min_{\boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}, \boldsymbol{\mathcal{P}}} \frac{1}{2}\|\boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}\|_F^2 + \lambda\sum_{i=1}^{n_t}\|\boldsymbol{\mathcal{G}}_i\|_{\mathrm{NSTMR}}$$

$$\text{s.t. } \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{G}}_i = R_i(\boldsymbol{\mathcal{P}}), \quad i = 1, 2, \ldots, n_t. \quad (5)$$

By introducing the multipliers $\boldsymbol{\mathcal{U}}$ and $\{\boldsymbol{\mathcal{V}}_i\}_{i=1}^{n_t}$, associated to the linear constraints $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{G}}_i = R_i(\boldsymbol{\mathcal{P}}),\ i = 1, 2, \ldots, n_t$, respectively, the augmented Lagrangian function for (5) is given as follows:

$$\mathcal{L} = \frac{1}{2}||\boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}||_F^2 + \lambda \sum_{i=1}^{n_t} ||\boldsymbol{\mathcal{G}}_i||_{\text{NSTMR}} + \frac{\beta_1}{2}||\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{U}}||_F^2 + \frac{\beta_2}{2}\sum_{i=1}^{n_t} ||\boldsymbol{\mathcal{G}}_i - R_i(\boldsymbol{\mathcal{P}}) + \boldsymbol{\mathcal{V}}_i||_F^2 \tag{6}$$

where β_1 and β_2 are penalty parameters.

The optimization problem is well structured because all the variables are divided into two groups $(\boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t})$ and $\boldsymbol{\mathcal{P}}$. The minimization problem can be separated into two smaller subproblems, so that two groups of variables $(\boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t})$ and $\boldsymbol{\mathcal{P}}$ can be minimized in alternating order before updating multipliers.

In Step 1, we update $\boldsymbol{\mathcal{X}}$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}$ in the following subproblem:

$$\min_{\boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}} \frac{1}{2}||\boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}||_F^2 + \lambda \sum_{i=1}^{n_t} ||\boldsymbol{\mathcal{G}}_i||_{\text{NSTMR}} + \frac{\beta_1}{2}||\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{U}}||_F^2 + \frac{\beta_2}{2}\sum_{i=1}^{n_t} ||\boldsymbol{\mathcal{G}}_i - R_i(\boldsymbol{\mathcal{P}}) + \boldsymbol{\mathcal{V}}_i||_F^2. \tag{7}$$

The optimal solutions of $\boldsymbol{\mathcal{X}}$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}$ can be calculated separately since they are decoupled. The $\boldsymbol{\mathcal{X}}$-subproblem

$$\min_{\boldsymbol{\mathcal{X}}} \frac{1}{2}||\boldsymbol{\mathcal{D}} \odot (\boldsymbol{\mathcal{B}} \circledast \boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{Y}}||_F^2 + \frac{\beta_1}{2}||\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{U}}||_F^2 \tag{8}$$

can be breaked up into the following subproblems equivalently in the vector form:

$$\min_{\mathbf{x}_i} \frac{1}{2}\|\mathbf{D}_i\mathbf{B}_i\mathbf{x}_i - \mathbf{y}_i\|_F^2 + \frac{\beta_1}{2}\|\mathbf{x}_i - \mathbf{p}_i + \mathbf{u}_i\|_F^2, i = 1, \ldots, b \tag{9}$$

where b is the number of bands, $\mathbf{x}_1, \ldots, \mathbf{x}_b \in \mathbb{R}^{MN\times 1}$ are the target high spatial resolution bands obtained by vectorizing each band image, i.e., $\mathbf{x}_i = \texttt{vec}(\boldsymbol{\mathcal{X}}^{(i)})$, $\mathbf{y}_1, \ldots, \mathbf{y}_b \in \mathbb{R}^{MN\times 1}$ are the upsampled low spatial resolution bands, i.e., $\mathbf{y}_i = \texttt{vec}(\boldsymbol{\mathcal{Y}}^{(i)})$, $\mathbf{B}_1, \ldots, \mathbf{B}_b \in \mathbb{R}^{MN\times MN}$ are matrices, which are equal to frontal slices of each block tensor $\boldsymbol{\mathcal{B}}_i \in \mathbb{R}^{M\times M\times N}$ of $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{M\times MB\times N}$, i.e., $\mathbf{B}_i = \texttt{bcirc}(\boldsymbol{\mathcal{B}}_i)$, representing the PSF of the sensor, $\mathbf{D}_1, \ldots, \mathbf{D}_B \in \mathbb{R}^{MN\times MN}$ are diagonal downsampling matrices, which are equal to the vectorization of frontal slices of $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{M\times N\times N}$, i.e., $\texttt{diag}(\mathbf{D_i}) = \texttt{vec}(\boldsymbol{\mathcal{D}}^{(\mathbf{i})})$, and $\mathbf{u}_i$ and $\mathbf{p}_i$ are obtained by vectorizing each frontal slices of $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{P}}$, respectively, i.e., $\mathbf{u}_i = \texttt{vec}(\boldsymbol{\mathcal{U}}^{(i)})$ and $\mathbf{p}_i = \texttt{vec}(\boldsymbol{\mathcal{P}}^{(i)})$. In essence, (8) and (9) are the same for modeling and degradation process. They are the different forms for solving $\boldsymbol{\mathcal{X}}$. In the ADMM framework, the variables $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{G}}$ are solved as the unitary block, which is tensor-based form and describes the low rankness of the data. As we can see, variables $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{G}}$ are decoupled, $\boldsymbol{\mathcal{X}}$ subproblem is reformulated as (9), which is tensor-based form. In order to solve the problem efficiently, the problem is equal to a vector-based form which is (9).

Each $\mathbf{x}_i$-subproblem has the following closed-form solution:

$$\mathbf{x}_i = (\mathbf{B}_i^H\mathbf{D}_i^H\mathbf{D}_i\mathbf{B}_i + \beta_1\mathbf{I}_{N\times N})^{-1}(\mathbf{B}_i^H\mathbf{D}_i^H\mathbf{y}_i + \beta_1(\mathbf{p}_i - \mathbf{u}_i)) \tag{10}$$

which can be calculated exactly and efficiently via fast Fourier transforms (FFTs) [43], [44]. The cost of updating $\mathbf{x}$ is $O(BMN\log MN)$.

For the $\boldsymbol{\mathcal{G}}_i$-subproblem

$$\min_{\boldsymbol{\mathcal{G}}_i} \frac{\beta_2}{2}\|\boldsymbol{\mathcal{G}}_i - R_i(\boldsymbol{\mathcal{P}}) + \boldsymbol{\mathcal{V}}_i\|_F^2 + \lambda\|\boldsymbol{\mathcal{G}}_i\|_{\text{NSTMR}},\ i = 1, \ldots, n_t$$

according to [39], [45], it has the closed-form solution

$$\boldsymbol{\mathcal{G}}_i = \texttt{ifft}\left(\texttt{Fold}\left(\begin{bmatrix}\mathbf{U}_i^1 \mathrm{D}_{\frac{\lambda}{\beta_2},\varepsilon}(\boldsymbol{\Sigma}_i^1)\mathbf{V}_i^{1T}\\ \mathbf{U}_i^2 \mathrm{D}_{\frac{\lambda}{\beta_2},\varepsilon}(\boldsymbol{\Sigma}_i^2)\mathbf{V}_i^{2T}\\ \vdots \\ \mathbf{U}_i^b \mathrm{D}_{\frac{\lambda}{\beta_2},\varepsilon}(\boldsymbol{\Sigma}_i^b)\mathbf{V}_i^{bT}\end{bmatrix}\right)\right), i = 1, \ldots, n_t \tag{11}$$

where $\mathbf{U}_i^k\boldsymbol{\Sigma}_i^k\mathbf{V}_i^{kT}, k = 1, \ldots, B$ is the singular value decomposition of $(\overline{R_i(\boldsymbol{\mathcal{P}})}^{(k)} - \overline{\boldsymbol{\mathcal{V}}}_i^{(k)})$ and $\mathrm{D}_{\frac{\lambda}{\beta_2},\varepsilon}(\boldsymbol{\Sigma}_i^k)$ is thresholding operator defined as

$$\mathrm{D}_{\frac{\lambda}{\beta_2},\varepsilon}(x) = \begin{cases} 0 & \text{if} \quad c_2 \le 0 \\ \text{sign}(x)\left(\frac{c_1 + \sqrt{c_2}}{2}\right) & \text{if} \quad c_2 > 0 \end{cases}$$

with $c_1 = |x| - \varepsilon, c_2 = (c_1)^2 - 4(\frac{\lambda}{\beta_2} - \varepsilon|x|)$. The cost of updating all $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{n_t}$ is $O(n_t n_s \min\{WB^2, B^2W\})$.

In Step 2, we update $\boldsymbol{\mathcal{P}}$ in the following subproblem:

$$\min_{\boldsymbol{\mathcal{P}}} \frac{\beta_1}{2}\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{U}}\|_F^2 + \frac{\beta_2}{2}\sum_{i=1}^{n_t}\|\boldsymbol{\mathcal{G}}_i - R_i(\boldsymbol{\mathcal{P}}) + \boldsymbol{\mathcal{V}}_i\|_F^2$$

which has the closed-form solution

$$\boldsymbol{\mathcal{P}} = \left(\beta_1(\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{U}}) + \beta_2\sum_{i=1}^{n_t} R_i^H(\boldsymbol{\mathcal{G}}_i + \boldsymbol{\mathcal{V}}_i)\right)./(\beta_1\mathbf{1} + \beta_2\boldsymbol{\mathcal{W}}). \tag{12}$$

Here, $\boldsymbol{\mathcal{W}} := \sum_{i=1}^{n_t}\boldsymbol{\mathcal{W}}_i$ and $\boldsymbol{\mathcal{W}}_i$ computes the number of times that each pixel of $\boldsymbol{\mathcal{X}}$ occurs in $\boldsymbol{\mathcal{G}}_i$. The cost of updating $\boldsymbol{\mathcal{P}}$ is $O(MNB)$. In the last step, we update the multipliers

$$\boldsymbol{\mathcal{V}}_i := \boldsymbol{\mathcal{V}}_i + \boldsymbol{\mathcal{G}}_i - R_i(\boldsymbol{\mathcal{P}}),\ i = 1, 2, \ldots, n_t \tag{13}$$

and

$$\boldsymbol{\mathcal{U}} := \boldsymbol{\mathcal{U}} + \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{P}}. \tag{14}$$

The solving algorithm is summarized in Algorithm 1.

In the following part, we discuss the convergence of the proposed Algorithm 1. Before that, we give a brief review about the framework and convergence of the ADMM in nonconvex nonsmooth optimization problem [46]. Wang *et al.* [46] considered the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = 0 \end{aligned} \tag{15}$$

Algorithm 1: The ADMM Algorithm for Solving NSTMR.

Input:
1: $\boldsymbol{\mathcal{Y}}$ - observed image
2: $\boldsymbol{\mathcal{B}}$ - blurring tensor
3: $\boldsymbol{\mathcal{D}}$ - mask tensor
4: β_1, β_2 - penalty parameters
5: λ - regularization parameter;
Initialization: $\boldsymbol{\mathcal{X}}^0, \boldsymbol{\mathcal{G}}_i^0, \boldsymbol{\mathcal{P}}^0, \boldsymbol{\mathcal{V}}_i^0, \boldsymbol{\mathcal{U}}^0$
6: **while** not converged **do**
7: Update $\boldsymbol{\mathcal{X}}$ via Eq. (10);
8: Update $\boldsymbol{\mathcal{G}}_i$ via Eq. (11);
9: Update $\boldsymbol{\mathcal{P}}$ via Eq. (12);
10: Update the multiplier $\boldsymbol{\mathcal{V}}_i$ via Eq. (13);
11: Update the multiplier $\boldsymbol{\mathcal{U}}$ via Eq. (14);
12: **end while**
Output: The estimation of $\boldsymbol{\mathcal{X}}$

where $f : \mathbb{R}^I \to \mathbb{R}^L$ is continuous, proper, possibly nonsmooth, and nonconvex, $g : \mathbb{R}^J \to \mathbb{R}^L$ is proper, differentiable and possibly nonconvex, $\mathbf{x} \in \mathbb{R}^I$ and $\mathbf{y} \in \mathbb{R}^J$ are variables with the corresponding coefficient matrices $\mathbf{A} \in \mathbb{R}^{L\times I}$ and $\mathbf{B} \in \mathbb{R}^{L\times J}$, respectively. By introducing the auxiliary multiplier $\mathbf{z} \in \mathbb{R}^L$, we have the augmented Lagrangian function $\mathcal{L}_\beta$ as

$$\mathcal{L}_\beta(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{z}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}\rangle + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}\|_2^2$$

where $\beta > 0$ is a penalty parameter. ADMM solves (15) iteratively as following:

$$\begin{cases} \mathbf{x}^{\mathbf{t+1}} = \arg\min_{\mathbf{x}} \mathcal{L}_\beta(\mathbf{x}, \mathbf{y}^{\mathbf{t}}, \mathbf{z}^{\mathbf{t}}) \\ \mathbf{y}^{t+1} = \arg\min_{\mathbf{y}} \mathcal{L}_\beta(\mathbf{x}^{t+1}, \mathbf{y}, \mathbf{z}^t) \\ \mathbf{z}^{t+1} = \mathbf{z}^t + \beta(\mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1}). \end{cases} \tag{16}$$

The following Lemma 1 gives the convergence of ADMM in nonconvex nonsmooth optimization.

Lemma 1: Suppose that the following assumptions A1–A5 hold, then for any sufficiently large β and starting from any $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0)$, the sequence generated by (16) has at least one limit point, and each limit point is a stationary point of $\mathcal{L}_\eta$.

A1 (*coercivity*): The objective function $f(\mathbf{x}) + g(\mathbf{y})$ is coercive over the nonempty feasible set $\mathcal{D} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{I+J} : \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = 0\}$.

A2 (*feasibility*): $Im(\mathbf{A}) \subseteq Im(\mathbf{B})$, where $Im(\cdot)$ denotes the image of a matrix.

A3 (*Lipschitz subminimization paths*):

(a) For any $\mathbf{x}$, there exists a Lipschitz continuous map $h : Im(\mathbf{B}) \to \mathbb{R}^L$ obeying $h(\mathbf{u}) = \arg\min_{\mathbf{y}}\{f(\mathbf{x}) + g(\mathbf{y}) : \mathbf{B}\mathbf{y} = \mathbf{u}\}$.

(b) For any $\mathbf{y}$, there exists a Lipschitz continuous map $k : Im(\mathbf{A}) \to \mathbb{R}^L$ obeying $k(\mathbf{u}) = \arg\min_{\mathbf{x}}\{f(\mathbf{x}) + g(\mathbf{y}) : \mathbf{A}\mathbf{x} = \mathbf{u}\}$.

A4 (*objective-f regularity*): f is lower semicontinuous or $\sup\{\|d\| : x \in S, d \in \partial f(x)\}$ is bound for any bound set S.

A5 (*objective-g regularity*): g is Lipschitz differentiable.

Next, we present the convergence result of the proposed Algorithm 1.

Theorem 1: For any sufficiently large β, Algorithm 1 generates a sequence $(\mathcal{X}^t, \{\mathcal{G}_i^t\}, \mathcal{P}^t, \mathcal{U}^t, \{\mathcal{V}_i^t\})$ that converges to the stationary point of the augmented Lagrangian function $\mathcal{L}_\eta$.

Proof: We first reformulate (6) in the matrix-vector multiplication form as

$$\arg\min_{\mathcal{Z},\mathcal{P}} \quad f(\mathcal{Z}) + g(\mathcal{P})$$
$$\text{s.t.} \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & \cdots & 0 \\ 0 & 0 & \vdots & \vdots \\ 0 & 0 & \cdots & I \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{n_t} \end{pmatrix} - \begin{pmatrix} I \\ R_1 \\ \vdots \\ R_{n_t} \end{pmatrix} \mathbf{p} = \mathbf{0}$$

where $\mathcal{Z} = [\mathcal{X}; \mathcal{G}_1; \cdots; \mathcal{G}_{n_t}]$, $f(\mathcal{Z}) = \frac{1}{2}\|\mathcal{D} \odot (\mathcal{B} \circledast \mathcal{X}) - \mathcal{Y}\|_F^2 + \lambda \sum_i \|\mathcal{G}_i\|_{\text{NSTMR}}$, $g(\mathcal{P}) = 0$, I denotes the identity matrix, R_i is the matrix that extracts the pixel form $\mathcal{P}$ (since the $\mathcal{R}_i$ is the linear operator), and $\mathbf{x}$, $\{\mathbf{g}_i\}_{i=1}^{n_t}$, and $\mathbf{p}$ denote the vectorization of $\mathcal{X}$, $\{\mathcal{G}_i\}_{i=1}^{n_t}$, and $\mathcal{P}$, respectively. It is clear that our model fits the framework of (15).

Now we check the assumptions A1–A5. A1 holds because of the coercivity of $f(\mathcal{Z}) + g(\mathcal{P})$. A2 and A3 hold because the coefficient matrix of $\mathbf{p}$ and the coefficient matrix of the column vector composed by $\mathbf{x}$ and $\{\mathbf{g}_i\}_{i=1}^{n_t}$ are full column rank. A4 holds because $f(\mathcal{Z})$ is lower semicontinuous. A5 holds because $g(\mathcal{P})$ is Lipschitz differentiable.

IV. Experimental Results

In this section, we evaluate the performance of the proposed algorithm on simulated data with S2 parameters and real S2 images. All experiments are performed using Windows 10 and MATLAB Version R2017a running on a desktop with an Intel Core i9-7900 K CPU at 3.30 GHz with 64 GB of memory.

Quantitative indices: For simulated data, the signal-to-reconstruction error (SRE), the spectral angle mapper (SAM), and structural similarity (SSIM) index are used for quantitative performance evaluations. The SRE of each band is defined as

$$\text{SRE} := 10\log_{10}\frac{||\boldsymbol{\mathcal{X}}^{(k)}||_F^2}{||\boldsymbol{\mathcal{X}}^{(k)} - \hat{\boldsymbol{\mathcal{X}}}^{(k)}||_F^2}$$

measured in dB, where $\hat{\boldsymbol{\mathcal{X}}}^{(k)}$ and $\boldsymbol{\mathcal{X}}^{(k)}$ are, the kth band of the estimated image and the ground truth image, respectively. The average SRE (aSRE) is defined as

$$\text{aSRE} := \frac{1}{(b_{20} + b_{60})} \sum_{i=1}^{(b_{20}+b_{60})} \text{SRE}_\text{i}$$

where b_{20} and b_{60} are the numbers of 20 m and 60 m bands, and SRE_i is the SRE of ith 20 m or 60 m band. The SAM [47] is computed as

$$\text{SAM} := \frac{1}{mn(b_{20} + b_{60})} \sum_{i=1}^{mn} \arccos \frac{\langle \mathbf{X}_{:i}, \hat{\mathbf{X}}_{:i}\rangle}{||\mathbf{X}_{:i}||^2 ||\hat{\mathbf{X}}_{:i}||^2}$$

where $\hat{\mathbf{X}}_{i:}$ and $\mathbf{X}_{i:}$ are the spectral values of 20 m and 60 m bands located at the pixel i, of the estimated image $\hat{\mathbf{X}}$ and the ground truth image $\mathbf{X}$.

The SSIM [48] is computed as

$$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) := \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + c_2)} \tag{17}$$

where $\mu_{\mathbf{x}}$ and $\mu_{\hat{\mathbf{x}}}$ are the average of groundtruth and the estimated image, $\sigma_{\mathbf{x}}^2, \sigma_{\hat{\mathbf{x}}}^2$ are the variance, $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$ is the covariance. The SSIM is obtained by using the MATLAB command `ssim`.

The SAM and SSIM is computed only for the super-resolved bands and then is averaged over all pixels of the image. Generally speaking, better results are reflected by higher SRE values and lower SAM values. The stopping criterion of ADMM is the relative change of the successive iterations to be less than a specified tolerance: $\|\boldsymbol{\mathcal{X}}^k - \boldsymbol{\mathcal{X}}^{k-1}\|_F / \|\boldsymbol{\mathcal{X}}^{k-1}\|_F \leq 1 \times 10^{-3}$.

Compared methods: We compare the proposed NSTMR method with ATPRK [11], SupReME [16], MuSA [17], and S2Sharp [18] methods. In the abovementioned methods, except for the first method, which belongs to pansharpening algorithms, all other methods fall into the model-based class. The reason why we did not compare with deep learning (DL) methods was that DL-based methods depends on neural network, whose output results largely subject to distribution of training dataset. In our experiments compared with DSen2, the dataset we use was not identically distributed with the DSen2's so that the performance is not easily compared by SRE and SAM. Most importantly, the proposed method is model based so we compared with methods of the same type.

Simulated data: For the simulation of S2 images, we use both simulated and real data. Simulated S2 images are generated from airborne hyperspectral images. There are three simulated datasets. The first two datasets are based on the HYDICE image of Washington DC Mall[1] and Terrain[2] with 2.8 m spatial resolution. The third dataset we use is the airborne prism experiment (APEX) [49] image of Baden[3] with 1.8 m spatial resolution. We use the following strategies to generate simulated images. First, we created the ground truth (GT) images, which is the super-resolved S2 images with 10 m resolution at all 12 bands. For this purpose, we selected twelve bands with the same wavelength as S2 satellite to generate the MS images. We lowpass filtered the abovementioned MS images and then downsampled the blurred MS image with a factor of $d_1 = 3$ to approximately obtain a spatial resolution of 10 m. The reason for this operation is that other satellites do not necessarily have a ground sample distance of 10 meters in the same wavelength and factor d_1 depends on the ratio of ground sample distance between S2 and the satellite generating the simulated images. Then, we blurred the abovementioned GT to obtain the images with spatial features at 20 m and 60 m and downsampled the blurred images with a factor of $d_2 = 2$ and $d_3 = 6$, obtaining the observed images. Take the Washington DC as an example, the size of satellite image is $280 \times 307 \times 191$, after selecting bands, lowpass filtering, and downsampling, the GT is $90 \times 96 \times 12$. Next, we blurred the GT and downsampled the result to obtain simulated 20 m (downsampling factor $d_2 = 2$) and simulated 60 m bands (downsampling factor $d_3 = 6$), obtaining the observed LR bands with the size of 15×16 (60 m), 90×96 (10 m), 45×48 (20 m), respectively.

The Gaussian blur kernels differ from each band with 10 m, 20 m, and 60 m. These kernels are computed from the calibrated modulation transfer function supplied by ESA.[4] These simulated S2 images are generated by blurring, downsampling and adding Gaussian noise such that the SNR is 40 dB. The size of Washingtong DC, Terrain, and APEX is $90 \times 96 \times 12$, $96 \times 96 \times 12$, and $114 \times 114 \times 12$, respectively.

For all three simulated datasets, the pixel values of each band are normalized scaled to the interval [0, 1]. The similar patches are found in the 10 m bands for these bands are ground truth and has more accurate spatial information. The SRE and SAM values of the estimated results by different methods on simulated data sets (Washington DC, Terrain, and APEX) are summarized in Table III. The bold font in this table denotes the best results. Apart from quantitive assessments, the visual effect of the residual image between recovered results and ground truth is also shown.

For the Washington DC image, the false-color images created with bands at 60 m (B1 and B11) and 20 m (B6) and the residual images at B1, B11, and B6 are shown in Fig. 3. In terms of the aSRE, it is easily seen that NSTMR performs the best, just with being a step behind S2sharp at B7, B8a, and B12 SRE. NSTMR not only delivers the best quantitive results but also outperforms the other methods at B1 and B6 for visual inspection on residual images. Most of the areas on the residual appear blue, which means the super-resolved image is closer to the GT.

For the Terrain image, the false-color images created with bands at 60 m (B1) and 20 m (B5 and B9) and residual image at B1, B5, and B9 are shown in Fig. 4. For the APEX image, the false-color images created with bands at 20 m (B5, B6, and B7) and residual image at B1, B7, and B9 are shown in Fig. 5. The residual images also show a decent improvement in visual effect. We can observe that the NSTMR method is better than the other methods since the residual images are closer to zero than other methods. For Terrain, the aSRE of NSTMR is the highest but the SRE at B1 a little less than S2Sharp. For APEX, aSRE of NSTMR is still the best but B1, B11, and B12 have slightly weaker results simply because these channels are of LR (60 m resolution) that the tensor multirank prior has a piece of less information about image pixels. Apart from the SRE, SAM is also reported in Table III. Here, for all three datasets, NSTMR also obtains the best result. NSTMR takes nonlocal spatial similarity and global spectral rudundancy into consideration so that it can deal with the S2 resolution problem effectively.

[1]Online. [Available]: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

[2]Online. [Available]: http://www.erdc.usace.army.mil/Media/Fact-Sheets/Fact-Sheet-Article-View/Article/610433/hypercube/

[3]Online. [Available]: http://www.apex-esa.org/content/free-data-cubes

[4]Online. [Available]: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_L1C_Data_Quality_Report

TABLE III
QUANTITIVE METRICS VALUES OF THE ESTIMATED RESULTS BY DIFFERENT ALGORITHMS ON SIMULATED DATASETS

Method	SRE								aSRE	SAM	SSIM	Time/Seconds
	B1	B5	B6	B7	B8a	B9	B11	B12				
Washington DC												
SupReMe	25.3112	31.0122	31.4326	32.1220	32.5455	23.3025	21.8230	18.8647	27.0517	2.3903	0.9700	0.6972
ATPRK	*	33.7361	32.8149	32.9793	32.7294	*	25.7546	23.2904	*	*	*	59.3909
MuSA	24.4736	27.8396	28.7723	28.1407	28.2222	25.0755	20.5715	18.0928	25.1485	2.5443	0.9800	87.8932
S2Sharp	25.0604	33.9656	35.8418	**36.7460**	**38.7878**	22.6184	23.0912	**23.3125**	29.9280	2.4692	0.9865	2.0018
NSTMR	**33.2047**	**36.2259**	**36.1371**	36.3370	37.0575	**31.5404**	**25.9400**	23.2688	**32.4639**	**1.3928**	**0.9887**	277.3383
Terrain												
SupReMe	22.1084	30.8999	31.8757	34.1488	35.7179	23.3498	24.1852	24.5096	28.3494	2.0560	0.9601	0.9117
ATPRK	*	33.5574	33.0843	32.9368	32.7859	*	27.5039	27.3637	*	*	*	60.5709
MuSA	29.3874	30.4232	30.2447	29.3983	29.5351	28.5039	24.1066	23.2421	28.1052	1.9431	0.9881	108.9950
S2Sharp	**29.5679**	29.1396	30.1327	34.0357	36.0061	29.9248	27.0796	27.2897	30.3970	1.3290	0.9893	5.5523
NSTMR	29.3845	**37.6998**	**37.8837**	**38.2669**	**38.5779**	**31.4993**	**28.8008**	**28.7166**	**33.8537**	**1.2174**	**0.9952**	223.4361
APEX												
SupReMe	18.5195	24.9552	25.3929	28.4696	31.4845	18.4084	14.8874	11.3419	21.6824	6.1018	**0.9737**	2.5050
ATPRK	*	**27.8107**	29.3474	31.9634	32.9172	*	**16.5317**	**13.0474**	*	*	*	60.6986
MuSA	**19.4365**	23.5539	24.8531	26.5852	28.0203	24.9665	13.2948	10.0710	21.3476	5.4061	0.9597	219.9476
S2Sharp	17.6436	26.7919	27.2818	30.3898	35.6363	22.1084	15.1809	11.9489	23.3727	4.7066	0.9595	6.1630
NSTMR	18.6472	26.2333	**29.7782**	**33.5253**	**35.7237**	**25.9477**	15.7967	12.3245	**24.7471**	**4.2765**	0.9652	614.0096

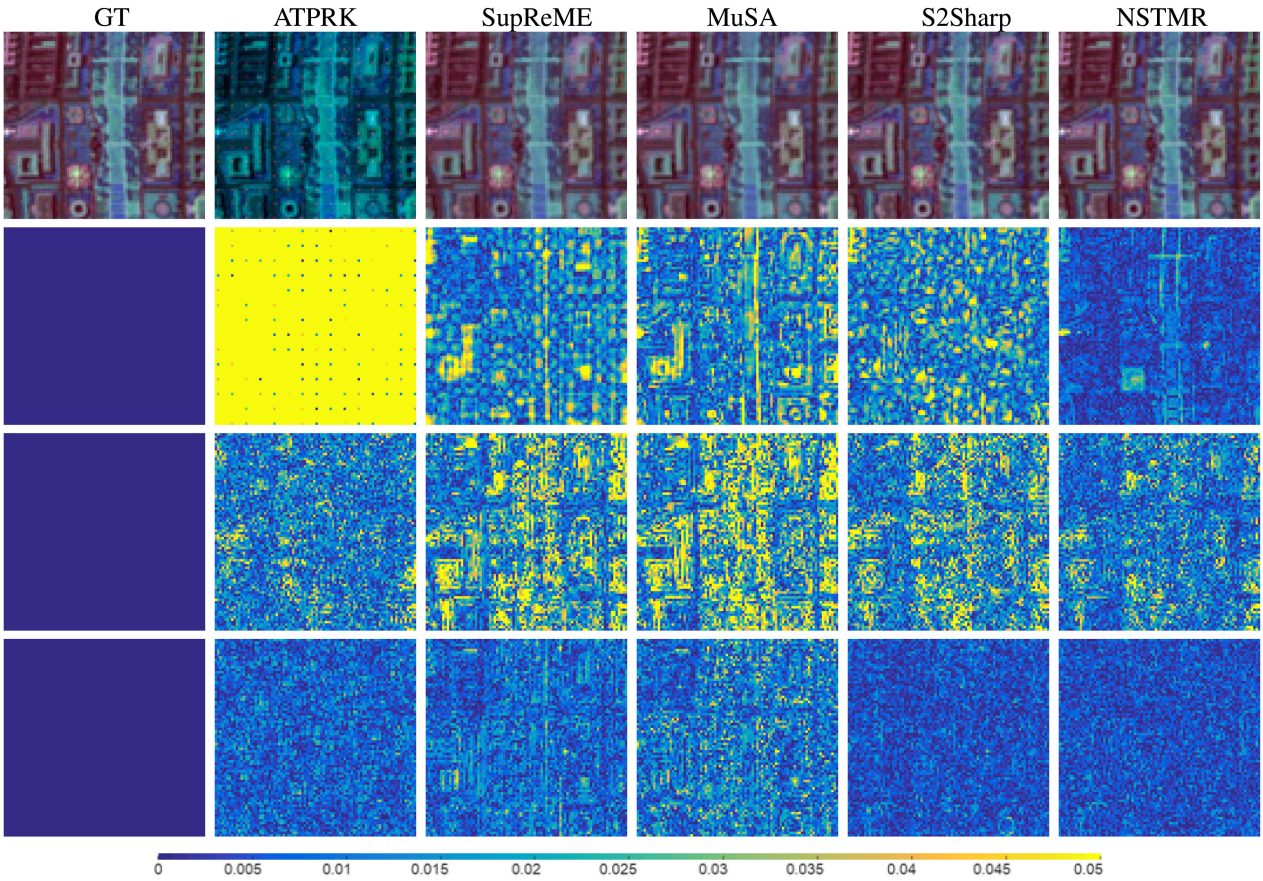

Fig. 3. Top: the false-color images created with bands B1, B11, and B6. Bottom: the residual image of the estimated results at B1, B6, and B11 by different methods for Washington DC. From left to right: the GT, results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

In terms of SSIM, the results in Washingtong DC and Terrain from NSTMR are 0.9887 and 0.9952, respectively, indicating a better recovery result compared with other methods, which was consistent with the trend of SRE and aSRE. At the same time, the SSIM on APEX is slightly inferior to SupReME, and in line with the trend of SRE being slightly inferior to other methods at B5, B11 and B12. The reason is that the low rankness of image APEX cannot be captured by the proposed model, because the content of image is not very evenly distributed buildings, i.e., the image spatial redundancy is not very good. As for processing time, it can be observed that MuSA and NSTMR have much longer processing time than other methods,

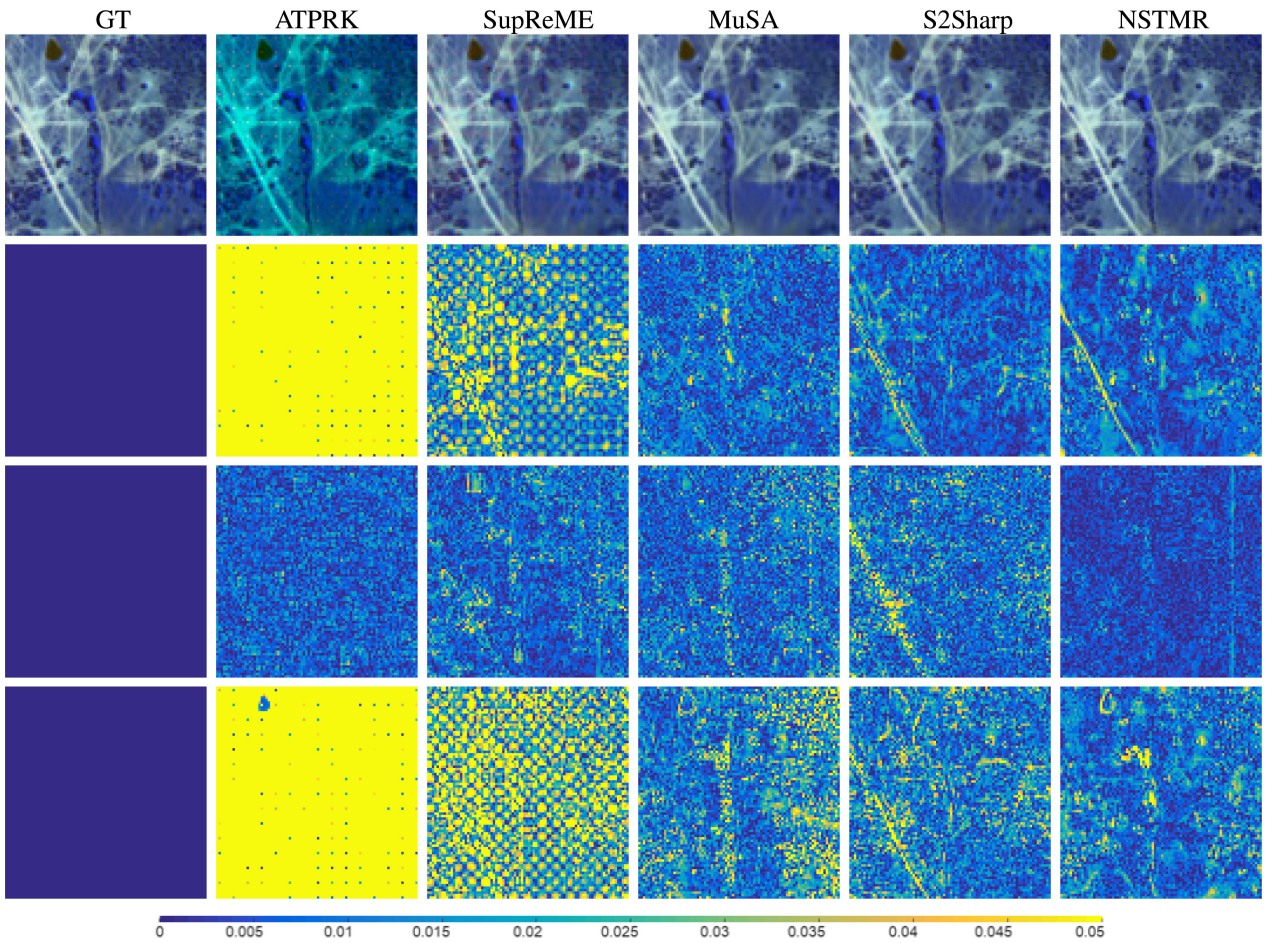

Fig. 4. Top: the false-color images created with bands B1, B5, and B9. Bottom: the residual image of the estimated results at B1, B5, and B9 by different methods for Terrain. From left to right: the GT, results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

due to the large amount of time spent searching for nonlocal similar patches.

Real data: For real datasets, we use real S2 images of scenes Verona, Malmo, Treviso,[5] and RealData4.[6] For the first three data, we select an area with a spatial extent of 180×180. For the last data, we select an bigger image with the size of 300×300. The pixel values of each band are also normalized scaled to the interval [0, 1] in real data. Due to the unreliability of blind measures, we only use the visual effect to give intuitive evaluations of different algorithms. For each data, we select an area with a spatial extent of 20×20 (30×20 for RealData4) and zoom it out in the lower right for a contrast effect. The false-color images created with bands at 20 m and 60 m are shown in Figs. 6 –9. In terms of visual inspection, the NSTMR method can match the S2Sharp and other competing methods. From bottom right corner of the image in Figs. 6 and 8 in this response, we can see that the SupReME and NSTMR provided more details and texture, showing the geometric details in the scene better. In Fig. 7, the details of the enlarged images by different methods are not very similar so it is difficult to estimate, which is better. In Fig. 9, the results by NSTMR and S2Sharp are very similar. When the data are large, the running time of MuSA and NSTMR is much longer than other algorithms since this two methods need to search for the nonlocal patches. Both methods use block matching 3-D stategies to process similar patches. As can be seen from the Fig. 9, there are many stripes with alternating rows and columns in the result estimated by MuSA, indicating that the parameters of this method may not optimal. For parameters on the real datasets, we used the best parameters from the simulated data with the same size. In order to get the optimal parameter, the more metrics should be introduced but there is no specific quantitative metrics to evaluate on the real set so that parameter tunning is a difficult task. However, for NSTMR, from the perspective of vision, the details can be compared with S2Sharp, therefore, NSTMR is also a good method.

[5]Online. [Available]: https://earthexplorer.usgs.gov/

[6]Online. [Available]: https://scihub.copernicus.eu/dhus/

V. Discussion

A. Convergence

In this section, we show the numberical convergence of Algorithm 1. In Fig. 10, we demonstrate the relative error curves with respect to iteration number in recovering of Washington DC, Terrain, and APEX. From the figure, we can observe that the

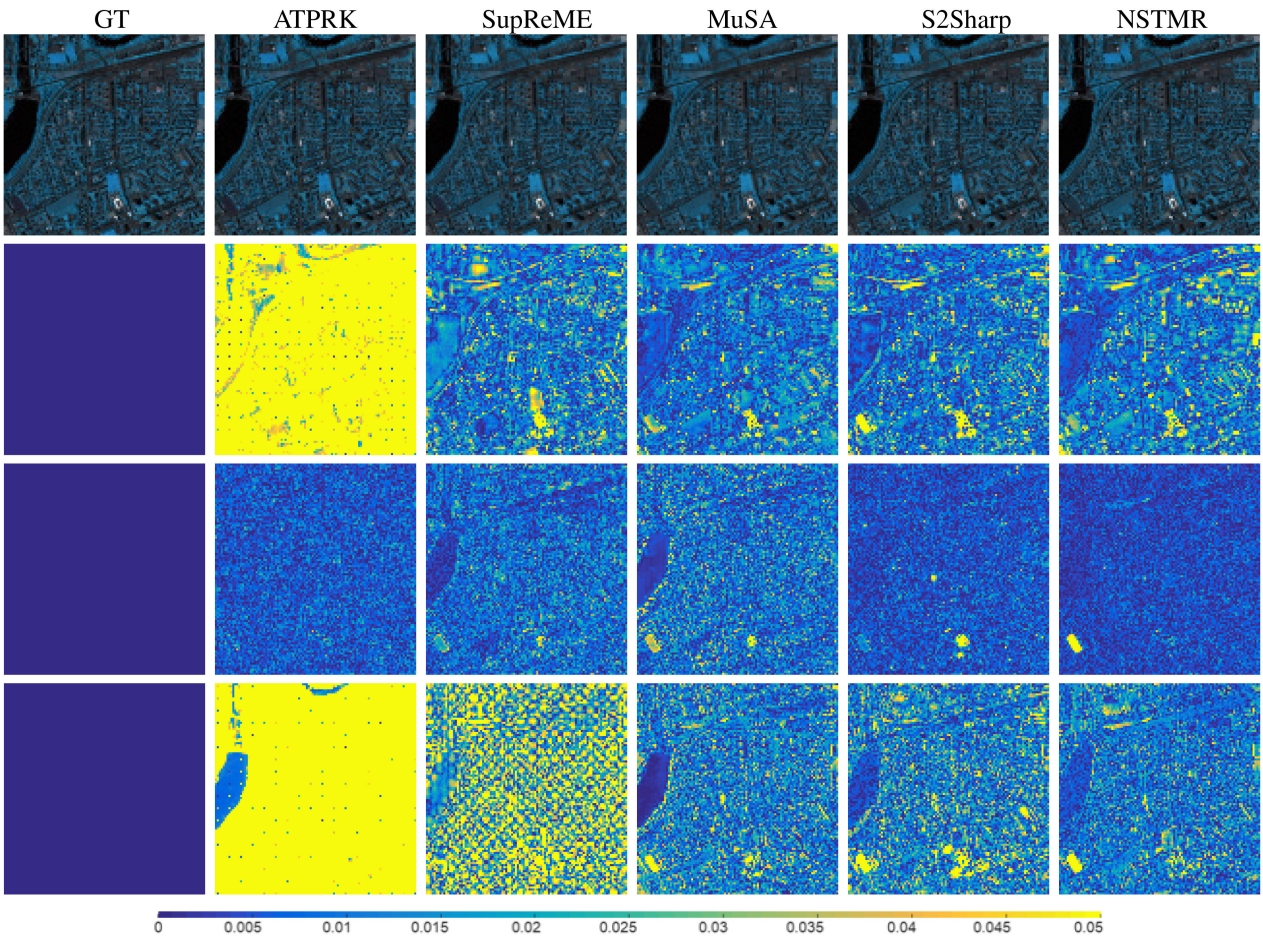

Fig. 5. Top: the false-color images created with bands B5, B6, and B7. Bottom: the residual image of the estimated results at B1, B7, and B9 by different methods for APEX. From left to right: the GT, results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

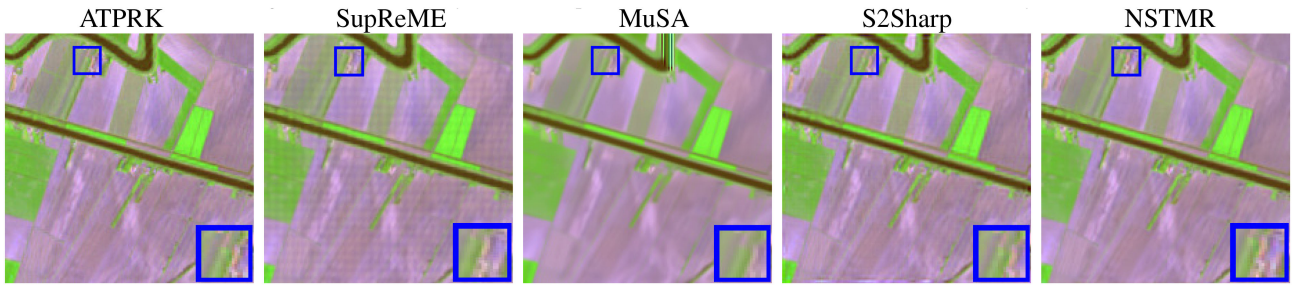

Fig. 6. False-color images created with bands at 20 m (B5, B6, and B12) by different methods for Treviso. From left to right: the estimated results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

algorithm stops the iteration quickly, which shows the numerical convergence of Algorithm 1.

B. Parameter Analysis

In this section, we analyze the effect of parameter λ, β_1, and β_2. The ATPRK main parameter are the size of the PSF. The SupReME main parameter is the regularization parameter λ. The MuSA main parameters are the regularization parameter λ and noise standard deviation τ. The S2Sharp main parameter is the regularization parameter λ. For other methods, we use grid parameter tuning. Their parameter selection range is $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and default number provided by the author. We select the parameters when aSRE is maxium. All parameters of competing methods are fine-tuned to achieve their best performance. Note that, fine tuning of parameters for different datasets may yield better results, but we unify the parameter selection to illustrate the robustness of the proposed

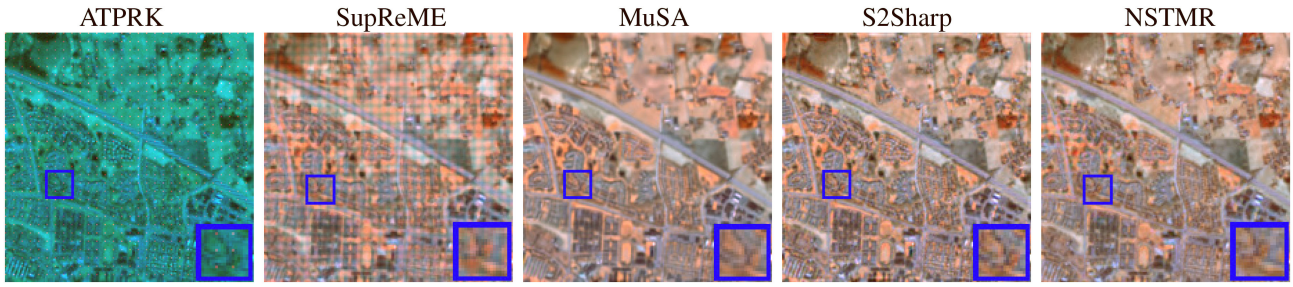

Fig. 7. False-color images created with bands at 60 m (B9) and 20 m (B11 and B12) by different methods for Malmo. From left to right: the estimated results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

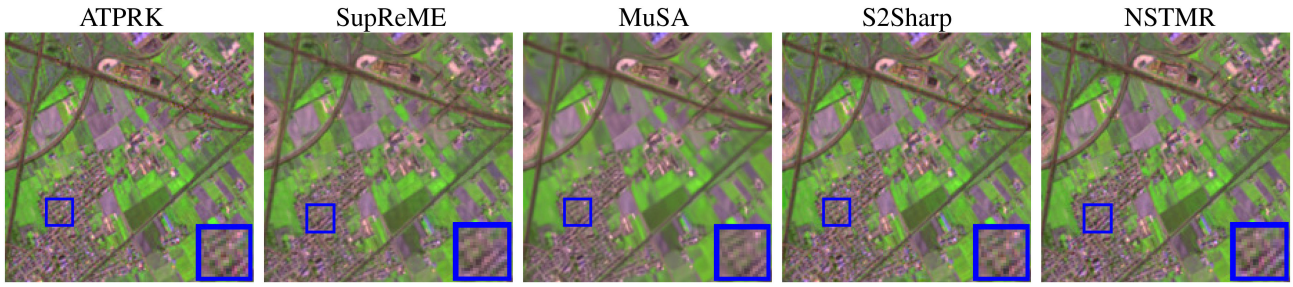

Fig. 8. False-color images created with bands at 20 m (B5, B6, and B12) by different methods for Verona. From left to right: the estimated results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

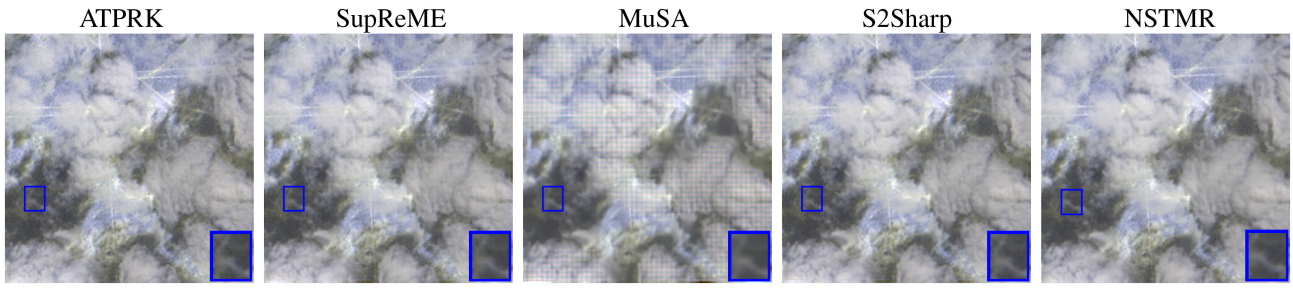

Fig. 9. False-color images created with bands at 20 m (B5, B6, and B12) by different methods for RealData4. From left to right: the estimated results by ATPRK, SupReME, MuSA, S2Sharp, and NSTMR, respectively.

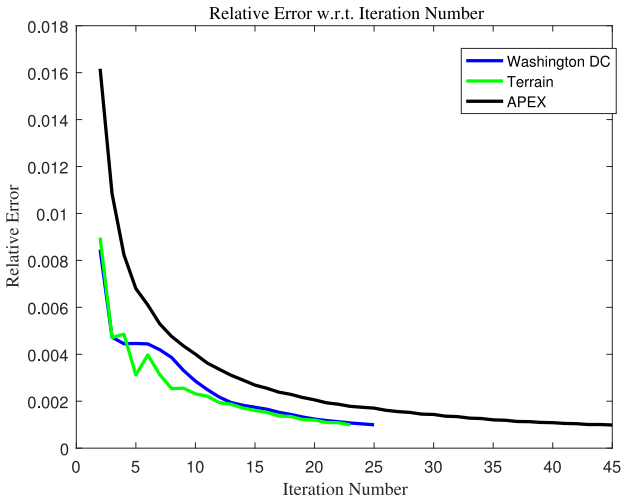

Fig. 10. Relative error curves with respect to iteration number on Washington DC, Terrain, and APEX.

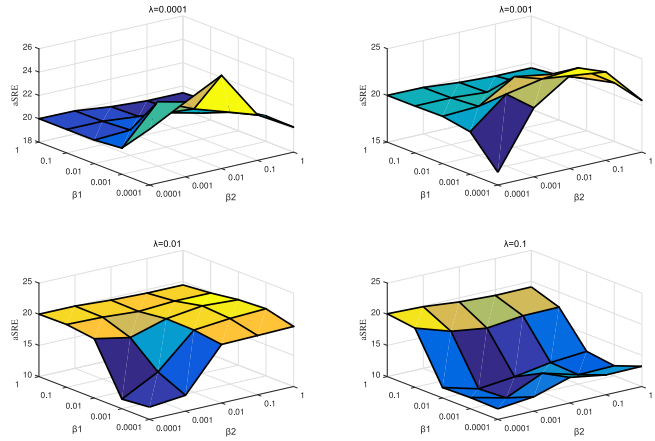

Fig. 11. The aSRE with respect to β_1, β_2, and λ.

method. Taking the APEX as an example, in the process of

fine tuning, we select $\beta_1, \beta_2, \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ and $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ to estimate the optimal parameters. In Fig. 11, we show the aSRE values with respect to β_1, β_2, and λ. We can see that as λ rises from 10^{-4} to 10^{-1}, the range of aSRE decreases and NSTMR delivers the relatively good performance when λ is between 10^{-3} and 10^{-4}. In both cases, when β_1 is smaller and β_2 are between 0.01 and 0.1, the aSRE is much higher. As smaller β_1 and β_2 make the algorithm achieve convergence in more steps, β_1 and β_2 cannot get any smaller values to achieve faster convergence. Under the tradeoff of the speediness of algorithm convergence and performance, $\lambda = 0.001$, $\beta_1 = 0.1$, and $\beta_2 = 0.0001$ are the best parameters.

VI. Conclusion

To enhance the resolution of S2 images, we proposed a tensor-based nonlocal low-multirank regularized model by taking full advantage of the nonlocal spatial-spectral redundancy of the S2 image. The efficient alternating direction method of multipliers was developed to solve the proposed model and its convergence is theoretically guaranteed. Extensive experiments on simulated and real data demonstrated state-of-the-art performance of the proposed model both quantitatively and visually. It is remarkable that although upscaling 60 m bands to 10 m spatial resolution seems challenging, the proposed method NSTMR can obtain good results by taking full advantages of the low-rankness depicted by nonlocal similarity. In the future, we will design some new points to utilize the correlation of different bands.

References

[1] M. Berger *et al.*, "The sentinel missions-new opportunities for science," *Remote Sens. Environ.*, vol. 120, pp. 1–276, 2012.

[2] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "Esa's sentinel missions in support of earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, 2012.

[3] Z. Malenovský *et al.*, "Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, 2012.

[4] M. Drusch *et al.*, "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[5] M. Claverie *et al.*, "The harmonized landsat and Sentinel-2 surface reflectance data set," *Remote Sens. Environ.*, vol. 219, pp. 145–161, 2018.

[6] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[7] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, 2007.

[8] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2012, pp. 552–599.

[9] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li, "Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 354.

[10] A. Vaiopoulos and K. Karantzalos, "Pansharpening on the narrow VNIR and SWIR spectral bands of sentinel-2," *ISPRS-Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. XLI-B7, pp. 723–730, 2016.

[11] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, 2016.

[12] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 305–319, 2018.

[13] F. Palsson, J. Sveinsson, and M. Ulfarsson, "Sentinel-2 image fusion using a deep residual network," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1290.

[14] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Single sensor image fusion using a deep convolutional generative adversarial network," in *Proc. 9th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2018, pp. 1–5.

[15] N. Brodu, "Super-resolving multiresolution images with band-independant geometry of multispectral pixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4610–4617, Aug. 2017.

[16] C. Lanaras, J. Bioucas-Dias, E. Baltsavias, and K. Schindler, "Super-resolution of multispectral multiresolution images from a single sensor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1505–1513.

[17] C. Paris, J. Bioucas-Dias, and L. Bruzzone, "A novel sharpening approach for superresolving multiresolution optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1545–1560, Mar. 2019.

[18] M. O. Ulfarsson, F. Palsson, M. Dalla Mura, and J. R. Sveinsson, "Sentinel-2 sharpening using a reduced-rank method," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6408–6420, Sep. 2019.

[19] C.-H. Lin and J. M. Bioucas-Dias, "An explicit and scene-adapted definition of convex self-similarity prior with application to unsupervised Sentinel-2 super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3352–3365, May 2020.

[20] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[21] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1672–1682, Jun. 2006.

[22] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 60–65.

[23] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[24] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[25] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.

[26] K. Wang, Y. Wang, X.-L. Zhao, J. C.-W. Chan, Z. Xu, and D. Meng, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor decomposition and spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7654–7671, Nov. 2020.

[27] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[28] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[29] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5344–5353.

[30] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *J. Comput. Appl. Math.*, vol. 363, pp. 124–144, 2020.

[31] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, and T.-Z. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Appl. Math. Comput.*, vol. 367, 2020, Art. no. 124783.

[32] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[33] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.

[34] Y.-Y. Liu, X.-L. Zhao, Y.-B. Zheng, T.-H. Ma, and H. Zhang, "Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045169.

[35] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral Super-Resolution With Coupled Tucker Approximation: Recoverability

and SVD-Based Algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 931–946, Jan. 2020, doi: 10.1109/TSP.2020.2965305.
[36] J. L. Wang,731 T. Z. Huang, X. L. Zhao, T. X. Jiang, and M. K. Ng,732 "Multi-dimensional visual data completion via low-rank tensor representation under coupled transform," *IEEE Trans. Image Process.*, vol. 30, pp. 3581–3596, Mar. 2021.
[37] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
[38] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3842–3849.
[39] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1692–1700.
[40] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
[41] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[43] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for l1-l2 problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, Aug. 2016.
[44] Q. Wei, N. Dobigeon, J.-Y. Tourneret, J. Bioucas-Dias, and S. Godsill, "R-fuse: Robust fast fusion of multiband images based on solving a sylvester equation," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1632–1636, Nov. 2016.
[45] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye, "A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 37–45.
[46] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, 2019.
[47] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *Summaries Tgird Annu. JPL Airborne Geosci. Workshop*, vol. 1, pp. 147–149, 1992.
[48] Z. Wang, A. C.H. R. BovikSheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[49] M. E. Schaepman *et al.*, "Advanced radiometry measurements and earth science applications with the airborne prism experiment (APEX)," *Remote Sens. Environ.*, vol. 158, pp. 207–219, 2015.

Xuan-Qi Wang received the B.S. degree in Network Engineering from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2021.

Teng-Yu Ji received the B.S. and Ph.D. degrees both in Mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2012 and 2018, respectively.

From 2016 to 2017, he was a Visiting Researcher with the Technical University of Munich, Germany. Since 2018, he has been an Assistant Professor with the Northwestern Polytechnical University, Xi'an, China. His research interests include tensor decomposition and applications, including tensor completion and remotely sensed image reconstruction.