Nonlocal Tensor-Based Sparse Hyperspectral Unmixing

Jie Huang, Member, IEEE, Ting-Zhu Huang, Member, IEEE, Xi-Le Zhao, and Liang-Jian Deng, Member, IEEE

Abstract—Sparse unmixing is an important technique for analyzing and processing hyperspectral images (HSIs). Simultaneously exploiting spatial correlation and sparsity improves substantially abundance estimation accuracy. In this article, we propose to exploit nonlocal spectral information in the HSI for the sparse unmixing problem. Specifically, we first group similar patches in the HSI, and then unmix each group by imposing simultaneous a low-rank constraint and joint sparsity in the corresponding third-order abundance tensor. To this end, we build an unmixing model with a mixed regularization term consisting of the sum of the weighted tensor trace norm and the weighted tensor $\ell_2,1$-norm of the abundance tensor. The proposed model is solved under the alternating direction method of multipliers framework. We term the developed algorithm as the nonlocal tensor-based sparse unmixing algorithm. The effectiveness of the proposed algorithm is illustrated in experiments with both simulated and real hyperspectral data sets.

Index Terms—Hyperspectral unmixing, joint sparsity, low-rank, nonlocal similarity, tensor.

I. INTRODUCTION

SPECTRAL unmixing of hyperspectral images (HSIs) has received much attention in recent years [1], [2]. Its basic task is to identify the spectral signatures of the materials (endmembers) and also the fractions (abundances) of these endmembers in each mixed HSI pixel. In the literature, a plethora of geometrical and statistical methods have been put forward to identifying the spectral signatures of endmembers, known as endmembers' extraction; see [3]–[7] and references therein. On the other hand, nonnegative matrix factorization-based blind hyperspectral algorithms have been well studied for the task [8]–[10]. In addition, another attractive research topic is to assume the spectral signatures of endmembers available and belonging to a predefined set usually called endmembers' dictionary, then diverse abundance estimation algorithms have been proposed in a semisupervised fashion [11]–[14].

The mixing process in numerous state-of-the-art unmixing algorithms is either characterized as linear [2] or nonlinear [15]. Linear mixture model (LMM) assumes that the spectral signature of each pixel in HSI is a linear combination of the spectral signatures of the endmembers, weighted by their corresponding abundances. Due to its simplicity and tractability, LMM has been widely adopted by diverse spectral unmixing algorithms (see [1], [2] and reference therein). For the remainder of this article, we will focus on the linear model in a semisupervised way.

Incorporating a priori information improves the abundance estimation accuracy. The sparsity assumption plays a key role in many semisupervised algorithms. Its adoption is justified by the fact that only a few of spectral signatures participate in the linear mixture process of each pixel, particularly compared with large-scale available dictionaries [11]. Practically, sparsity is imposed on abundances via the standard $\ell_1$-norm regularization [11] and its variants [8], [16]–[19]. Collaborative (or joint) sparsity scheme provides a refinement by simultaneously imposing the sparsity on all pixel in the scene [12], [20]–[22]. Exploiting spatial correlation in sparse unmixing framework has offered stimulating results. Typically, the total variation (TV) is included in sparse unmixing to promote piecewise smooth in each abundance map [13], [23]–[26]. The low-rank property is another structural characteristic that has been recently adopted for abundances [22], [27], [28]. It transfers the spatial correlation among pixels’ spectral signatures into linear dependence among their corresponding abundance vectors. The resulting abundance matrix therefore admits a low-rank behavior and its rank is usually approximated by the weighted nuclear norm in practice. Nonlocal sparse unmixing presents an alternative way of modeling the spatial correlation. Nonlocal means method is incorporated into the TV regularization to exploit the similar patterns and structures in abundance maps [29], [30]. As an improvement of nonlocal means TV spatial consideration, an adaptive nonlocal Euclidean medians approach is proposed to better utilize the spatial information by means of filtering [31]. In addition, the nonlocal means method has been used in a collaborative sparse unmixing framework to suppress the abundance estimation error [32]. Recently, the nonlocal similarity of abundance maps is employed in unmixing problem by first grouping nonlocal similar abundance patches and then imposing a low-rank constraint [33], [34]. These nonlocal methods in [29], [31]–[34] mainly focus on exploring the nonlocal spatial information in abundance maps. Note that the abundances are
estimated iteratively, so the nonlocal spatial information is correspondingly updated as soon as the latest estimation of abundances is acquired.

In this article, we introduce a novel idea for exploring the nonlocal spatial information and spectral correlation. Departing from the usual paradigm, we explore the nonlocal spatial similarity in the HSI for hyperspectral unmixing. That said, we first group similar patches in the HSI and each group is stacked to be a third-order tensor. The technique of exploiting the spatial nonlocal similarity in HSIs has been well studied for HSI denoising [35]–[39]. It helps to improve the denoising performance, in which repetitive patches can provide more complementary information especially under heavy noise. Following this technique, we incorporate nonlocal spatial similarity and spectral correlation of HSIs into the unmixing process in an attempt to achieve better abundance estimation results. To the best of our knowledge, it is new for hyperspectral unmixing. Particularly, there is no need to recalculate the nonlocal spatial information in abundance maps at each iteration round. Then, we unmix each group by transferring the spatial and spectral correlation in each group into the low-rank property and the joint sparsity in the corresponding third-order abundance tensor. The low-rank and group into the low-rank property and the joint sparsity in the

The proposed algorithm is compared with state-of-the-art unmixing techniques, Section IV and a real-data experiment in Section V. Finally, Section VI gives some concluding remarks.

II. SPARSE UNMIXING MODEL

Let $Y_o \in \mathbb{R}^{n_1 \times n_2 \times J}$ denote the observed HSI with $n_1 \times n_2$ pixels and $J$ spectral bands. Then, the third-order tensor $Y_o$ is rearranged to a matrix $Y \in \mathbb{R}^J \times N$ with $N = n_1 \times n_2$ [34], [40]. Let $A \in \mathbb{R}^{J \times m}$ denote the dictionary with $m$ spectral signatures. The classic LMM can be described as

$$Y = AX + E$$  \hspace{1cm} (1)

where $X \in \mathbb{R}^{m \times N}$ is the fractional abundance matrix whose columns correspond with the abundance vectors of $N$ pixels in $Y$, and $E \in \mathbb{R}^J \times N$ is an independent and identically distributed (i.i.d.) zero-mean Gaussian noise matrix. With large and available $A$, sparse unmixing aims to find a relatively small number of spectral signatures to linearly mix every pixel in $Y$. Exploiting a prior information helps to improve the estimation accuracy. The so-called abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC)

$$X \geq 0, \quad 1^T X = 1^T$$

respectively, are widely considered for physical meaning [44]. Here, $X \geq 0$ denote that each element is nonnegative and $1$ is a column vector of 1’s. Nevertheless, we relax the ASC to focus

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>spatial width of HSI</td>
</tr>
<tr>
<td>$n_2$</td>
<td>spatial height of HSI</td>
</tr>
<tr>
<td>$J$</td>
<td>spectral bands of HSI</td>
</tr>
<tr>
<td>$K$</td>
<td>number of spectral bands</td>
</tr>
<tr>
<td>$n_i$</td>
<td>number of endmembers</td>
</tr>
<tr>
<td>$N$</td>
<td>number of pixels in HSI</td>
</tr>
<tr>
<td>$I$</td>
<td>number of pixels in a patch</td>
</tr>
<tr>
<td>$\mathbf{Y} \in \mathbb{R}^{J \times N}$</td>
<td>HSI matrix</td>
</tr>
<tr>
<td>$\mathbf{A} \in \mathbb{R}^{J \times m}$</td>
<td>spectral library</td>
</tr>
<tr>
<td>$\mathbf{X} \in \mathbb{R}^{m \times N}$</td>
<td>abundance matrix</td>
</tr>
<tr>
<td>$\mathbf{\mathbf{X}}_n$</td>
<td>tensor of similar HSI patches</td>
</tr>
<tr>
<td>$\mathbf{\mathbf{\mathbf{X}}}_m$</td>
<td>tensor of abundance</td>
</tr>
<tr>
<td>$\mathbf{\mathbf{\mathbf{Y}}}_{\mathbf{n}}$</td>
<td>mode-$n$ unfolding of $\mathbf{\mathbf{X}}$</td>
</tr>
<tr>
<td>$\mathbf{\mathbf{X}}_{i,1}$</td>
<td>$i$-th horizontal slice of $\mathbf{\mathbf{X}}$</td>
</tr>
<tr>
<td>$\mathbf{\mathbf{X}}_{i,2}$</td>
<td>$i$-th row of $\mathbf{\mathbf{X}}$</td>
</tr>
<tr>
<td>$|\mathbf{X}|_W$</td>
<td>weighted tensor trace norm</td>
</tr>
<tr>
<td>$|\mathbf{X}|_{2,1}$</td>
<td>weighted tensor $\ell_{2,1}$-norm</td>
</tr>
<tr>
<td>$|\mathbf{X}|_{z,2,1}$</td>
<td>weighted tensor $\ell_{2,1}$-norm</td>
</tr>
</tbody>
</table>

TABLE I

SOME NOTATIONS USED IN THIS ARTICLE
on the exploitation of structural characters of $X$, similarly as in [13], [12], [28], and [22]; see more details in [11].

To explore the sparsity of each abundance vector, the classic sparse regression model for hyperspectral unmixing problem is as follows:

$$\min_X \frac{1}{2}\|Y - AX\|_F^2 + \lambda \|X\|_{1,1}$$

s.t. $X \geq 0$ (2)

where $\|X\|_{1,1} = \sum_{i=1}^m \sum_{j=1}^N |x_{i,j}|$ is the $\ell_1$-norm of $X$, $x_{i,j}$ denotes the $(i, j)$th element of $X$, and $\lambda \geq 0$ is the regularization parameter.

A sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) algorithm is introduced to solve the above model in [11]. In addition, the $\ell_p$-norm-based regularization has been considered to promote sparseness in the abundance and also in other HSI processing problems, see [16], [17], [45] and reference therein.

With the assumption that all pixels in the data set share the same active set of endmembers, the collaborative sparse regression framework is

$$\min_X \frac{1}{2}\|Y - AX\|_F^2 + \lambda \|X\|_{2,p}$$

s.t. $X \geq 0$ (3)

where $p \in [0, 1]$ and $\lambda \geq 0$ is the regularization parameter. Denote $x^{[i]}$ be the $i$th row of $X$. Suppose $p = 1$, then $\|X\|_{2,1} = \sum_{i=1}^m \|x^{[i]}\|_2$ is the so-called $\ell_{2,1}$-norm of $X$. It follows a convex optimization problem and the corresponding algorithm is termed as collaborative SUnSAL (CLSUnSAL) in [12]. Consider instead $p = 0$ and $\|X\|_{2,0} = \sum_{i=1}^m 1(\|x^{[i]}\|_2 > 0)$ is the $\ell_{2,0}$-quasi-norm of $X$, it is shown in [21] that the $\ell_0$-quasi-norm regularization provides sparser and more accurate estimations. The corresponding algorithm is termed as the collaborative sparse hyperspectral unmixing algorithm using $\ell_0$-quasi norm (CSUnL0). In addition, a more general $\ell_{2,p}$-norm ($0 < p < 1$) regularization has been considered in [20] to promote sparsity along rows of the abundance matrix.

To exploit local collaborative sparsity property, a joint-sparse-blocks regression model has been proposed to encourage that similar adjacent pixels share the same support set of endmembers [22]. To this end, the abundance matrix is viewed as a block matrix

$$X = [X_1, \ldots, X_s]$$

(4)

where the column number of $X_j \in \mathbb{R}^{m \times d_j}$ satisfies $\sum_{j=1}^s d_j = N$ and the block number $s$ is a positive integer for $1 \leq s \leq N$. Then, each block $X_j$ is assumed joint-sparse. The so-called $\ell_{2,1}$-blocks regularized hyperspectral unmixing model is

$$\min_X \frac{1}{2}\|Y - AX\|_F^2 + \lambda \sum_{j=1}^s \|X_j\|_{2,1}$$

s.t. $X \geq 0$ (5)

where $\lambda \geq 0$ is the regularization parameter. Particularly, the model in (5) generalizes both the SUnSAL model by setting $s = N$ and the CLSUnSAL model by setting $s = 1$.

Recently, imposing the low-rank property on the abundance matrix provides a new perspective for exploiting spatial information [22], [27], [28]. Simultaneously imposing the sparsity and the low-rank representation on the abundance leads to an unmixing model as follows:

$$\min_X \frac{1}{2}\|Y - AX\|_F^2 + \lambda \|X\|_{Z,1} + \tau \|X\|_{w,s}$$

s.t. $X \geq 0$ (6)

where $\lambda$ and $\tau$ are nonnegative regularization parameters

$$\|X\|_{Z,1} = \sum_{i=1}^m \sum_{j=1}^N z_{i,j}|x_{i,j}|, \|X\|_{w,s} = \sum_{i=1}^r w_i \sigma_i$$

are the weighted $\ell_1$-norm and the weighted nuclear norm of $X$, respectively. $Z = [z_{i,j}] \in \mathbb{R}^{m \times N}$ and $w = [w_1, \ldots, w_r] \in \mathbb{R}^r$ are nonnegative weighting arrays, $\sigma_i$ is the $i$th singular value of $X$, and $r$ is the rank of $X$. The model in (6) can be solved by an algorithm termed as the alternating direction sparse and low-rank unmixing algorithm (ADSpLRU) [28]. Instead simultaneously imposing the joint-sparse-blocks structure and the low-rank representation on the fractional abundance, the unmixing model becomes

$$\min_X \frac{1}{2}\|Y - AX\|_F^2 + \lambda \sum_{j=1}^s \|X_j\|_{Z,j,2,1} + \tau \|X\|_{w,s}$$

s.t. $X \geq 0$ (7)

where $X_j$ is the $j$th column block of $X$ in (4), $\|X_j\|_{Z,j,2,1}$ is the weighted $\ell_{2,1}$-norm of $X_j$ defined as

$$\|X_j\|_{Z,j,2,1} = \sum_{i=1}^m z_{i,j} \|x^{[i]}_j\|_2$$

$z_j = [z_{1,j}, \ldots, z_{m,j}]^T \in \mathbb{R}^m$ is a nonnegative weighting vector for $X_j$, and $x^{[i]}_j$ is the $i$th row of $X_j$, for $i = 1, \ldots, m$ and $j = 1, \ldots, s$. The model in (7) is solved by the so-called joint-sparse-blocks and low-rank unmixing algorithm (JSpBLRU) and a reweighting strategy has been introduced to adaptively update $z_{i,j}$ at each iteration [22].

Finally, it is worth mentioning that tensor decomposition-based spectral unmixing has attracted considerable attention in recent years. A matrix-vector nonnegative tensor factorization (MV-NTF) has proposed for hyperspectral unmixing [46]. It approximates a third-order HSI tensor by sum of the outer products of an endmember and low-rank unmixing algorithm (ADSpLRU) [28]. Instead simultaneously imposing the joint-sparse-blocks structure and the low-rank representation on the fractional abundance, the unmixing model becomes
results. In the following, we propose a tensor-based unmixing algorithm that promotes the utilization of nonlocal spatial information and spectral correlation in HSIs in a different way in Section III. The advantages will be demonstrated later in Sections IV and V.

III. NL-TSU N ALGORITHM

In this section, we will first propose a nonlocal tensor-based unmixing optimization model and then derive an algorithm under the ADMM framework.

A. Proposed Unmixing Model

Unlike the authors in [29], [31]–[34] utilize the nonlocal spatial information in abundance maps, we consider directly exploiting the nonlocal spatial similarity in HSIs to further improve the abundance estimation performance. The process mainly consists of three stages: 1) group similar patches inside an HSI; 2) unmix each group by a novel tensor-based algorithm; and 3) aggregate to the abundance tensor of the HSI. The overall flowchart of the proposed algorithm is shown in Fig. 1. Specifically, at the first stage, we consider each key fullband patch (FBP), which is stacked by patches at the same location of the HSI over all bands, and there are many FBPs similar to the key FBP. We build a third-order tensor by stacking these similar FBPs, named group of similar HSI patches in Fig. 1. It is worth mentioning that utilizing the nonlocal spatial similarity has achieved state-of-the-art performance for the hyperspectral denoising problem [35]–[39]. Then, at the second stage, we unmix each group of similar HSI patches by encoding the nonlocal spatial similarity and spectral correlation in the corresponding abundance tensor. Finally, at the last stage, we aggregate the estimated groups of similar abundance patches to reconstruct the abundance tensor for the HSI. With the flowchart in mind, we study the second stage, that is, to unmix each group of similar HSI patches in the following.

Denote \( \mathbf{Y}^{(g)} \in \mathbb{R}^{I \times J \times K} \) as the third-order tensor grouped by similar nonlocal HSI patches, where \( I \) is the number of pixels, \( J \) is the number of spectral bands, and \( K \) is the number of similar patches. In the following, we use \( \mathbf{Y} \) to denote \( \mathbf{Y}^{(g)} \) for notation simplicity. Notice that the mode-2 of the third-order tensor \( \mathbf{Y} \) is the spectral mode. Then, the LMM for \( \mathbf{Y} \) is written as

\[
\mathbf{Y}(2) = \mathbf{A}\mathbf{X}(2) + \mathbf{E}(2)
\]

which is equivalent to, under the definition of the tensor mode-\( n \) multiplication

\[
\mathbf{Y} = \mathbf{X} \times_2 \mathbf{A} + \mathbf{E}
\]

where \( \mathbf{X} \) is the \( I \times m \times K \) fractional abundance tensor with \( m \) being the number of endmembers, \( \mathbf{A} \in \mathbb{R}^{J \times m} \) is the spectral library, and \( \mathbf{E} \in \mathbb{R}^{I \times J \times K} \) is the noise tensor.

In order to exploit the structure information of the abundance tensor \( \mathbf{X} \) stacking by similar abundance patches, we show the mode-\( n \) unfolding of \( \mathbf{X} \) in Fig. 2. From Fig. 2, we see that each mode-\( n \) unfolding of \( \mathbf{X} \) is low-rank. We remark that the low-rank property of the mode-\( n \) unfolding stems from the spatial correlation of pixels within a small size HSI patch for \( n = 1 \), from the high mutual coherence of the spectral libraries for \( n = 2 \), and from the nonlocal similarity of the HSI for \( n = 3 \). Therefore, to exploit the low-rankness characteristic of each mode-\( n \) unfolding, we define the following weighted tensor trace norm [50]:

\[
\|\mathbf{X}\|_{W^*,*} = \sum_{n=1}^{3} \|\mathbf{X}(n)\|_{W^*_n,*} = \sum_{n=1}^{3} \sum_{l=1}^{\text{rank}(\mathbf{X}(n))} w_{l,n} \sigma_{l,n}(\mathbf{X}(n))
\]
where $\sigma_{l,n}(X_{(n)})$ is the $l$th singular value of $X_{(n)}$ and $w_n = [w_{1,n}, \ldots, w_{\text{rank}(X_{(n)})}, \ldots]^T$ is a nonnegative weighting vector for $X_{(n)}$. Recall that the value of the weight $w_{l,n}$ is set to 1/3 for any $l$ and $n$ in [50]. Here, we propose to enhance the sparsity of the singular values of $X_{(n)}$ by assigning different weighting coefficients to different singular values, similarly as in [28], [36], and [22]. We will show the details in Section III-B. We note that the weighted nuclear norm has been used to characterize a global sparsity prior inside the tensor data in tensor completion problems [51].

We also see from Fig. 2 that each horizontal slice of the abundance tensor $\mathcal{X}$ admits a joint-sparse property and it therefore leads to a joint-sparse-blocks structure in the mode-2 unfolding matrix. This mainly due to the sparse assumption and the spatial nonlocal similarity in the abundance tensor. It is assumed that in sparse hyperspectral unmixing, only a small portion of the $m$ endmembers in the spectral library $A$ will be present in real HSIs. In other words, it is safe to assume that each abundance vector along the endmember mode in Fig. 2 has a sparse representation in terms of the potentially very large spectral library. Then, the spatial nonlocal similarity promotes that the corresponding abundance vectors of the similar pixels in the same horizontal slice share the same sparsity structure. It, therefore, follows a joint sparse structure in each horizontal slice of the abundance tensor $\mathcal{X}$. Clearly, it is a generalization of the matrix case.

To utilize the joint sparsity of $\mathcal{X}$, we define the following weighted tensor $\ell_{2,1}$-norm:

$$\|\mathcal{X}\|_{z_{2,1}} = \sum_{i=1}^l \|X_{i:} \|_{z_{2,1}} = \sum_{i=1}^l \sum_{k=1}^m z_{k,i} \|x_{i:}^{[k]}\|_2$$ (10)

where $x_{i:}^{[k]}$ is the $k$th row of $X_{i:}$ and $z_i = [z_{1,i}, \ldots, z_{m,i}]^T$ is the $i$th column of the nonnegative weighting matrix $Z \in \mathbb{R}^{m \times l}$. That said, we impose the joint sparsity on each horizontal slice of $\mathcal{X}$, i.e., $X_{i:} \in \mathbb{R}^{m \times K}$, via the weighted $\ell_{2,1}$-norm. Particularly, weighting coefficients $z_{k,i}$s are assigned differently to each row within each horizontal slice and they will be adaptively updated, similarly as in [22].

Simultaneously imposing the low-rank property via (9) and the joint sparsity via (10) on the abundance tensor $\mathcal{X}$, we obtain the following tensor-based optimization problem:

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \|_F^2 + \lambda_1 \|\mathcal{X}\|_{z_{2,1}} + \lambda_2 \|\mathcal{W}_{*,*}\|_{W_{*,*}}$$

s.t. $\mathcal{X} \succeq 0$ (11)

where $\lambda_1$ and $\lambda_2$ are nonnegative regularization parameters and $Z$ and $W$ are two nonnegative weighting matrices.

**B. NL-TSUn Algorithm**

To unmix each group of similar HSI patches $\mathcal{Y}$, we now solve the model in (11). To begin, we let $\Omega_x$ be the indicator function of a set $\Omega$, i.e., $\Omega_x(x) = 0$ if $x \in \Omega$ and $\Omega_x(x) = +\infty$ otherwise. Combining with (9) and (10), the model in (11) becomes

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{Y} - \mathcal{X} \|_F^2 + \lambda_1 \sum_{i=1}^l \|X_{i:} \|_{z_{i,2,1}}$$
$$+ \lambda_2 \sum_{n=1}^l \|X_{(n)}\|_{W_{*,*}} + \Omega_x(\mathcal{X})$$ (12)

Equivalently, we have

$$\min_{\mathcal{X}, \mathcal{M}_1, \ldots, \mathcal{M}_6} \frac{1}{2} \|\mathcal{Y} - \mathcal{M}_1 \|_F^2 + \lambda_1 \sum_{i=1}^l \|\mathcal{M}_{2,i:} \|_{z_{i,2,1}}$$
$$+ \lambda_2 (\|\mathcal{M}_{3} \|_{W_{*,*}} + \|\mathcal{M}_{4} \|_{W_{*,*}})$$
$$+ \Omega_x(\mathcal{M}_6)$$

s.t. $\mathcal{X} = \mathcal{M}_l, \quad l = 1, \ldots, 6$ (13)
where $M_{2,:i}$ is the $i$th horizontal slice of $\mathcal{M}_2$ and $M_{l(n)}$ is the mode-$n$ unfolding of $\mathcal{M}_l$, for $l = 1, \ldots, 6$ and $n = 1, 2, 3$.

We now solve (13) under the ADMM framework. Define

$$L_\mu(\mathbf{X},\mathbf{M}_1,\ldots,\mathbf{M}_6,\mathbf{D}_1,\ldots,\mathbf{D}_6) = \frac{1}{2}\|\mathbf{Y} - \mathbf{M}_1 \times_2 \mathbf{A}\|_F^2 + \lambda_1 \sum_{i=1}^l \|M_{2,:i}\|_{\ell_2,2,1} + \lambda_2 \|M_{3(1)}\|_{\ell_1} + \lambda_3 \|M_{4(2)}\|_{\ell_2} + \lambda_4 \|M_{5(3)}\|_{\ell_1,\ast} + \mu \|\mathbf{X} - \mathbf{M}_1 - \mathbf{D}_1\|_F^2,$$

where $\mu > 0$ is a penalty parameter. The ADMM framework is then derived

$$\begin{align*}
\mathcal{M}_1^{t+1} & = \text{argmin}_{\mathcal{M}_1} L_\mu(\mathbf{X}_t,\mathbf{M}_1,\ldots,\mathbf{M}_6;\mathbf{D}_1,\ldots,\mathbf{D}_6), \\
\mathbf{X}_t^{t+1} & = \text{argmin}_{\mathbf{X}_t} L_\mu(\mathbf{X}_t,\mathbf{M}_1^{t+1},\ldots,\mathbf{M}_6^{t+1};\mathbf{D}_1,\ldots,\mathbf{D}_6), \\
\mathbf{D}_l^{t+1} & = \mathbf{D}_l - \mathbf{X}_t^{t+1} + \mathbf{M}_l^{t+1}, l = 1, \ldots, 6.
\end{align*}$$

(15)

It is easy to see that the minimization problem about $\mathcal{M}_l$, $l = 1, \ldots, 6$, can be decoupled and so we can solve these subproblems separately.

1) For $\mathcal{M}_1$-subproblem, we have

$$\begin{align*}
\min_{\mathcal{M}_1} & \frac{1}{2}\|\mathbf{Y} - \mathbf{M}_1 \times_2 \mathbf{A}\|_F^2 + \mu \|\mathbf{X}_t - \mathbf{M}_1 - \mathbf{D}_1\|_F^2 \\
= & \min_{\mathcal{M}_1} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{A}\mathbf{M}_1\|_F^2 + \frac{\mu}{2}\|\mathbf{X}_t - \mathbf{M}_1 - \mathbf{D}_1\|_F^2. \\
\end{align*}$$

This is a least squares problem and we obtain

$$M_{1(2)}^{t+1} = (\mathbf{A}^T\mathbf{A} + \mu \mathbf{I})^{-1}\left(\mathbf{A}^T\mathbf{Y}_t + \mu (\mathbf{X}_t - \mathbf{D}_1^t)\right),$$

where $\mathbf{I}$ is an $m \times m$ identity matrix. With the fold operator, we have

$$\mathcal{M}_1^{t+1} = \text{fold}_2(M_{1(2)}^{t+1}).$$

(16)

2) For $\mathcal{M}_2$-subproblem, we have

$$\begin{align*}
\min_{\mathcal{M}_2} & \lambda_1 \sum_{i=1}^l \|M_{2,:i}\|_{\ell_2,2,1} + \mu \|\mathbf{X}_t - \mathbf{M}_2 - \mathbf{D}_2\|_F^2 \\
= & \min_{\mathcal{M}_2} \lambda_1 \sum_{i=1}^l \|\mathbf{M}_{2,:i}\|_{\ell_2,2,1} + \frac{\mu}{2}\|\mathbf{X}_t - \mathbf{M}_2 - \mathbf{D}_2\|_F^2.
\end{align*}$$

Clearly, we can equivalently decouple the above minimization problem to $l$ parts, that is, for $i$th horizontal slice of $\mathcal{M}_2$, we have

$$\begin{align*}
\min_{M_{2,:i}} & \lambda_1 \|\mathbf{M}_{2,:i}\|_{\ell_2,2,1} + \frac{\mu}{2}\|\mathbf{X}_t - \mathbf{M}_{2,:i} - \mathbf{D}_{2,:i}\|_F^2. \\
\end{align*}$$

(17)

for $i = 1, \ldots, l$. Recall from [52] that each subproblem admits a unique block solution, and the $k$th row of the unique block solution $M_{2,:i}$, i.e., $m_{2,:i}$ of (17), can be written explicitly as

$$m_{2,:i}^{t+1} = \text{vect-soft}_{\lambda_1}(\mathbf{X}_t^{t+1} - \mathbf{D}_{2,:i}^{t+1}),$$

for $k = 1, \ldots, m$, $i = 1, \ldots, l$, where $(\mathbf{X}_t)^{t+1}$ and $(\mathbf{D}_{2,:i}^{t+1})_{[k]}$ are the $k$th rows of $\mathbf{X}_t^t$ and $\mathbf{D}_{2,:i}^t$, respectively, and $\text{vect-soft}_{\lambda_1}(\cdot)$ is a nonlinear operator defined by

$$\text{vect-soft}_{\lambda_1}(\mathbf{x}) = \mathbf{x} + \max\{\|\mathbf{x}\|_2 - \alpha, 0\}$$

for $\forall \mathbf{x} \in \mathbb{R}^K$ and $\alpha > 0$. Here, similarly as in JSPBLRU, we employ a reweighting strategy to enhance sparsity along rows in $M_{2,:i}^{t+1}$ and for (18), set

$$z_{k,i} = \frac{1}{\|\mathbf{x}_i^t - \mathbf{d}_{2,:i}^t\|_2 + \epsilon}$$

(20)

where $\epsilon = 10^{-16}$ is a small constant added to avoid singularities. We note that the reweighting coefficient $z_{k,i}$ is an extension of weights for reweighted $\ell_1$ minimization [53]. It is shown in [22] that the abundance estimation accuracy has been substantially improved by using the reweighting coefficients. We will also show the effectiveness of $z_{k,i}$ in (20) in Section IV-B. Then, we get $\mathcal{M}_2^{t+1}$ by stacking all the horizontal slices $M_{2,:i}^{t+1}$, whose $k$th row is obtained from (18).

3) To solve the $\mathcal{M}_3$-subproblem, we first give some definitions. For any matrix $\mathbf{X}$, denote $r = \text{rank}(\mathbf{X})$ and let $\sigma_i$ denote the $i$th singular value of $\mathbf{X}$, for $i = 1, \ldots, r$. Denote elementwise $(\cdot)_+$ as $\max(\cdot, 0)$ and $\mathbf{X} = \text{UDiag}(\sigma_1, \ldots, \sigma_r)\mathbf{V}^T$ be the singular value decomposition (SVD) of $\mathbf{X}$. Define the weighted singular value thresholding operator $\text{SVT}_{\tilde{w},\beta}(-)$ on $\mathbf{X}$ as

$$\begin{align*}
\text{SVT}_{\tilde{w},\beta}(\mathbf{X}) & = \text{UDiag}((\sigma_1 - \beta \tilde{w}_1)_+, \ldots, (\sigma_r - \beta \tilde{w}_r)_+)\mathbf{V}^T \\
\end{align*}$$

with $\tilde{w} = [\tilde{w}_1, \ldots, \tilde{w}_r]$ being a nonnegative weighting vector. We note that the weighting coefficient $\tilde{w}_i = (1/\sigma_i + \epsilon)$ has designed to increase the punishment on smaller singular values and decrease the punishment on large singular values simultaneously. This technique has been widely used for many practical problems; see, e.g., [22], [28], [51], [54], [55]. Consider the $\mathcal{M}_3$-subproblem now. From (14) and (15), we get that

$$\begin{align*}
\min_{\mathcal{M}_3(1)} & \lambda_1 \|M_{3(1)}\|_{\ell_1} + \frac{\mu}{2}\|\mathbf{X}_t^t - \mathbf{M}_3(1) - \mathbf{D}_3(1)\|_F^2. \\
\end{align*}$$

(21)

Then, from [54] and [55], the closed-form solution of (21) is

$$M_{3(1)}^{t+1} = \text{SVT}_{\tilde{w}_1}(\mathbf{X}_t^t - \mathbf{D}_3(1)).$$

(22)

Here, similarly as in [54], [28], [55], [22], and [51], we set $\tilde{w}_1 = [w_{1,1}, \ldots, w_{1,r}]^T$ based on the singular values of the argument of the $\text{SVT}$ operator in (22). That is, define $\sigma_{i,1}$ is the $i$th singular values of $\mathbf{X}_t^t - \mathbf{D}_3(1)$ and $r_1$ is the rank of $\mathbf{X}_t^t - \mathbf{D}_3(1)$, then

$$w_{1,l} = \frac{1}{\sigma_{l,1} + \epsilon}, \quad l = 1, \ldots, r_1.$$

Then we have

$$\mathcal{M}_3^{t+1} = \text{fold}_1(M_{3(1)}^{t+1}).$$

(23)
4) Note that the minimization problems with respect to $\mathbf{M}_4$ and $\mathbf{M}_5$ are similar to that of $\mathbf{M}_3$. Thus, we obtain that

$$
\mathbf{M}_4^{t+1} = \text{fold}_2(\text{SVT}_{\mathbf{w}_2}^{\frac{\lambda}{2}}(\mathbf{X}_l^{t}(2) - \mathbf{D}_4^{t}(2)))
$$

$$
\mathbf{M}_5^{t+1} = \text{fold}_3(\text{SVT}_{\mathbf{w}_3}^{\frac{\lambda}{2}}(\mathbf{X}_l^{t}(3) - \mathbf{D}_5^{t}(3))).
$$

(24)

Here, we compute the weights $\mathbf{w}_2 = [w_{1,1}, \ldots, w_{r_2,2}]^T$ and $\mathbf{w}_3 = [w_{1,3}, \ldots, w_{r_3,3}]^T$ similarly as $\mathbf{w}_1$. That is

$$
\mathbf{w}_{l,u} = \frac{1}{\sigma_{l,u} + \epsilon}, \quad l = 1, \ldots, r_n
$$

(25)

in which $\sigma_{l,u}$ is the $l$th singular value of $\mathbf{X}_l^{t} - \mathbf{D}_4^{t}(2)(n)$, the argument of the SVT operator in (24), and $r_n$ is the rank of $\mathbf{X}_l^{t} - \mathbf{D}_4^{t}(2)(n)$ for $n = 1, 2, 3$.

5) For $\mathbf{M}_6$-subproblem, we have

$$
\mathbf{M}_6^{t+1} = \underset{\mathbf{M}_6}{\text{argmin}} \ t_{\mathbf{F}}(\mathbf{M}_6) + \frac{\mu}{2} \| \mathbf{X}^t - \mathbf{M}_6 - \mathbf{D}_6 \|_F^2.
$$

It is easy to obtain that

$$
\mathbf{M}_6^{t+1} = \max(\mathbf{X}^t - \mathbf{D}_6, 0).
$$

(26)

6) For $\mathbf{X}$-subproblem, it is to solve

$$
\min \frac{\mu}{2} \sum_{l=1}^6 \| \mathbf{X} - \mathbf{M}_l^{t+1} - \mathbf{D}_l \|_F^2,
$$

and thus

$$
\mathbf{X}^{t+1} = \frac{1}{6} \sum_{l=1}^6 (\mathbf{M}_l^{t+1} + \mathbf{D}_l).
$$

(27)

7) Finally, we update Lagrange multipliers $\mathbf{D}_l^{t+1}$, $l = 1, \ldots, 6$, according to (15).

To make it more clear, we summarize the proposed NL-TSUn in the following.

IV. EXPERIMENTS ON SIMULATED DATA

In this section, we illustrate the unmixing performance of the proposed NL-TSUn algorithm on simulated hyperspectral data sets. We will compare NL-TSUn with five state-of-the-art algorithms: SUnSAL [11], CLSUnSAL [12], SUnSAL-TV [13], ADSpLRU [28], and JSpBLRU [22].

Our tests were done by using MATLAB R2016a on a PC with 3.7 GHz Intel Core i7 and 64 GB memory. The floating-point precision is $10^{-16}$.

We use the signal-to-reconstruction error (SRE) and the root mean square error (RMSE) to evaluate the performance of unmixing results. They are defined as

$$
\text{SRE (dB)} = 10 \log_{10}\left(\frac{\sum_{i=1}^N \| \mathbf{x}_i \|_2^2}{\sum_{i=1}^N \| \mathbf{x}_i - \hat{\mathbf{x}}_i \|_2^2}\right)
$$

$$
\text{RMSE} = \sqrt{\frac{1}{mnN} \sum_{i=1}^N \| \hat{\mathbf{x}}_i - \mathbf{x}_i \|_2^2}
$$

where $N$ is the number of pixels, $m$ is the number of endmembers, and $\hat{\mathbf{x}}_i$ and $\mathbf{x}_i$ are estimated and exact abundance vectors of the $i$th pixel, respectively. Generally, the higher SRE values and lower RMSE values, the high quality of the unmixing results. For all algorithms, the regularization parameters are tuned to get maximum SRE values. We select optimal regularization parameters for all compared algorithms from the following sequence:

$$
\{0, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50\}
$$

and choose the optimal penalty parameter $\mu$ from $\{0.001, 0.01, 0.1, 1\}$.

We empirically set the HSI patch size $25 \times 25$, i.e., $I = 625$, and $K = 10$ in NL-TSUn for all the experiments. When the ADMM is applied to solve each group of similar HSI patches in NL-TSUn, we adopt the termination criterion similarly as in ADSpLRU and JSpBLRU. That is, define

$$
\mathbf{r'} = \left[\sum_{l=1}^6 \| \mathbf{X}^t - \mathbf{M}_l \|_F^2\right]^{\frac{1}{2}}, \quad \mathbf{d'} = \mu \left[\sum_{l=1}^6 \| \mathbf{M}_l^t - \mathbf{M}_l^{t-1} \|_F^2\right]^{\frac{1}{2}}
$$

to be the Frobenius norms of the primal residual and the dual residual at the $r$th ADMM iteration, respectively. Then, we stop the iteration if both the following two conditions hold:

$$
\mathbf{r'} \leq \zeta, \quad \mathbf{d'} \leq \zeta
$$

or when the number of iterations reached 500. Here, we set $\zeta = ((3m + J)IK)^{1/2}$. The positive relative tolerance $\zeta$ is empirically set to $5 \cdot 10^{-6}$. Similarly as the compared algorithms SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, and JSpBLRU, we initialize $\mathbf{X}^0 = \mathbf{M}_l^0 = \text{fold}_2((A^T A + \mu I)^{-1} A^T Y(2))$ and $\mathbf{D}_l^0 = 0$, for $l = 1, \ldots, 6$.  

---

Algorithm 1 Pseudocode of NL-TSUn

1. **Input:** $\mathcal{Y}_o$ and $A$
2. Construct groups of similar HSI patches $(\mathcal{Y}(g))_{g=1}^G$
3. for each group $\mathcal{Y}(g)$ do
4. //Solve the problem (13) by ADMM
5. Initialize: $\mathbf{X}^0$, $\mathbf{D}_l^0$, $\mathbf{M}_l^0$, $l = 1, \ldots, 6$, and set $t = 0$
6. Select parameters: $\lambda_1, \lambda_2, \mu$, maximum iterations
7. while not convergence do:
8. Compute $\mathbf{M}_4^{t+1}$ by (16)
9. Compute $\mathbf{M}_5^{t+1}$ by (18)
10. Compute $\mathbf{M}_3^{t+1}$ by (23)
11. Compute $\mathbf{M}_4^{t+1}$ and $\mathbf{M}_5^{t+1}$ by (24)
12. Compute $\mathbf{M}_6^{t+1}$ by (26)
13. Compute $\mathbf{X}^{t+1}$ by (27)
14. Update Lagrange multipliers $\mathbf{D}_l^{t+1}$ by (15)
15. $t = t + 1$
16. end while
17. Calculate $\mathbf{X}(g) = \mathbf{X}^{t+1}$
18. end for
19. Aggregate $(\mathbf{X}(g))_{g=1}^G$ to form the estimated abundance tensor $\hat{\mathbf{X}}$ for $\mathcal{Y}_o$
A. Comparison With Different Algorithms

We now compare NL-TSUn with different unmixing algorithms on the following two widely used simulated data cubes.

Example 1: The first data cube has 100 × 100 pixels with 100 bands per pixel. The spectral library \( A_1 \in \mathbb{R}^{100 \times 120} \) is a subset of a library of 262 spectral signatures with 100 spectral bands generally found on satellites, from the National Aeronautics and Space Administration (NASA) Johnson Space Center (JSC) Spacecraft Materials Spectral Database. According to LMM, nine endmember signatures are randomly selected from \( A_1 \) and the true corresponding abundance maps are shown in Fig. 3. We note that the exact abundance vector for each selected endmember is obtained by stacking the corresponding abundance map. After the above procedure, the true data cube is contaminated by white Gaussian i.i.d. noise with SNR = 20, 30, and 40 dB, respectively.

Example 2: The data cube contains 128 × 128 pixels with 224 spectral bands. We use the spectral library \( A_2 \in \mathbb{R}^{224 \times 240} \), a randomly selected subset of the U.S. Geological Survey (USGS) spectral library (splitlib06a), which comprises 498 spectral signatures with reflectance values measured in 224 spectral bands, distributed uniformly ranging from 0.4 to 2.5 \( \mu \text{m} \). We randomly choose five endmembers from \( A_2 \) to generate the true data cube by LMM. For illustrative purposes, corresponding true abundance maps are shown in Fig. 4. Then, the true data cube is contaminated by white Gaussian i.i.d. noise with the same SNR values adopted for Example 1.

For different unmixing algorithms, we show true and estimated abundance maps of endmembers #7 and #9 for Example 1 in Figs. 5 and 6, respectively. Similar results for endmembers #3 and #5 for Example 2 are shown in Figs. 7 and 8.

B. Reweighting Coefficient Choice and Effectiveness of the Proposed Tensor-Based Regularizers

In the proposed model (11), we impose the low-rank constraint via the weighted nuclear norm on each mode-\( n \) unfolding of \( \mathbf{X} \) and also impose the joint sparsity via the weighted \( \ell_{2,1} \)-norm on each horizontal slice of \( \mathbf{X} \). In the following, we will first show the efficiency of reweighting coefficients in the tensor-based regularizers and then illustrates the low-rank and joint-sparse characters of the estimated abundance tensors for groups of similar HSI patches.

1) Reweighting Coefficient Efficiency: Herein, we aspire to demonstrate the merits emerging from the utilization of the reweighting coefficients \( w_{l,n} \) in (25) and \( z_{k,i} \) in (20), which are used to enhance the sparsity of the singular values of the...
mode-$n$ unfolding $X^{(n)}$ and the joint sparsity of the horizontal slice $X_{i,:}$, respectively. To this end, four cases are considered as below.

**NL-TSUn**

1) with reweighting $w_{l,n}$ and $z_{k,i}$.

2) with only reweighting $z_{k,i}$, i.e., $z_{k,i}$ as in (20) but $w_{l,n} = 1$.

3) with only reweighting $w_{l,n}$, i.e., $w_{l,n}$ as in (25) but $z_{k,i} = 1$.

4) without reweighting $w_{l,n}$ and $z_{k,i}$, i.e., $z_{k,i} = w_{l,n} = 1$.

Table III lists the SRE (dB) and the RMSE values of the above four cases for Example 1 with SNR = 20 dB. It shows the efficiency of the reweighting coefficients $z_{k,i}$ and $w_{l,n}$ in NL-TSUn. Recall that in the second stage of NL-TSUn, we unmix each group of similar HSI patches under the ADMM framework. Here, for further illustration purpose, we study the abundance estimation accuracy when unmixing 12 groups of similar HSI patches. In particular, we show the SRE values obtained by NL-TSUn with different choices of weights $z_{k,i}$ and $w_{l,n}$ in Fig. 9. It is clear that NL-TSUn with reweighting both $w_{l,n}$ and $z_{k,i}$ gives optimal SRE values for all considered groups, which is in line with the observation from Table III. We also note that other simulated tests show a similar conclusion so we omit the details for space consideration.
2) Low-Rankness Regularizer: For the low-rank regularization term, we plot the singular values of the mode-$n$ unfoldings of a true third-order abundance tensor and its estimation by NL-TSUn for Example 1 with SNR = 20 dB in Fig. 10. As shown in the figure that the low-rank property is clear in the unfolding matrices, and moreover, the estimated singular values of each mode-$n$ unfolding are close to the exact ones. It shows the rationality and the effectiveness of the low-rankness constraint on the abundance tensor.

3) Joint-Sparsity Regularizer: Considering the joint-sparse regularization, we show eight horizontal slices of a true third-order abundance tensor and its estimations by NL-TSUn for Example 1 with SNR = 20 dB in Fig. 11. Recall that each group of similar abundance patches is a tensor of size $625 \times 120 \times 10$ with $25 \times 25$ pixels and 120 endmembers and ten similar patches. We note that we list eight horizontal
TABLE II
SRE (dB) AND RMSE BY DIFFERENT UNMIXING ALGORITHMS FOR EXAMPLES 1 AND 2

<table>
<thead>
<tr>
<th>SNR</th>
<th>Criteria</th>
<th>SUUnSAL</th>
<th>CLSUUnSAL</th>
<th>SUUnSAL-TV</th>
<th>ADSpLRU</th>
<th>JSpBLRU</th>
<th>NL-TSUn</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 dB</td>
<td>SRE</td>
<td>3.27</td>
<td>2.46</td>
<td>8.09</td>
<td>7.62</td>
<td>10.03</td>
<td>11.91</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0497</td>
<td>0.0546</td>
<td>0.0285</td>
<td>0.0302</td>
<td>0.0228</td>
<td>0.0183</td>
</tr>
<tr>
<td>30 dB</td>
<td>SRE</td>
<td>8.25</td>
<td>5.90</td>
<td>13.26</td>
<td>16.59</td>
<td>18.37</td>
<td>18.75</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0280</td>
<td>0.0368</td>
<td>0.0158</td>
<td>0.0107</td>
<td>0.0087</td>
<td>0.0084</td>
</tr>
<tr>
<td>40 dB</td>
<td>SRE</td>
<td>13.75</td>
<td>11.60</td>
<td>19.68</td>
<td>26.41</td>
<td>28.72</td>
<td>29.20</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0149</td>
<td>0.0191</td>
<td>0.0075</td>
<td>0.0035</td>
<td>0.0027</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

Example 2

<table>
<thead>
<tr>
<th>SNR</th>
<th>Criteria</th>
<th>SUUnSAL</th>
<th>CLSUUnSAL</th>
<th>SUUnSAL-TV</th>
<th>ADSpLRU</th>
<th>JSpBLRU</th>
<th>NL-TSUn</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 dB</td>
<td>SRE</td>
<td>4.92</td>
<td>5.64</td>
<td>12.13</td>
<td>13.02</td>
<td>14.52</td>
<td>15.52</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0262</td>
<td>0.0241</td>
<td>0.0114</td>
<td>0.0103</td>
<td>0.0087</td>
<td>0.0077</td>
</tr>
<tr>
<td>30 dB</td>
<td>SRE</td>
<td>11.90</td>
<td>14.59</td>
<td>19.31</td>
<td>21.20</td>
<td>22.93</td>
<td>24.11</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0117</td>
<td>0.0086</td>
<td>0.0050</td>
<td>0.0040</td>
<td>0.0033</td>
<td>0.0029</td>
</tr>
<tr>
<td>40 dB</td>
<td>SRE</td>
<td>19.07</td>
<td>22.49</td>
<td>28.24</td>
<td>31.66</td>
<td>32.73</td>
<td>33.44</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0051</td>
<td>0.0035</td>
<td>0.0018</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

TABLE III
SRE (dB) AND RMSE BY NL-TSUn WITH DIFFERENT WEIGHTING COEFFICIENTS FOR EXAMPLE 1 WITH SNR = 20 dB

<table>
<thead>
<tr>
<th>Criteria</th>
<th>reweighting $w_{l,n}$ and $z_{k,i}$</th>
<th>only reweighting $z_{k,i}$</th>
<th>only reweighting $w_{l,n}$</th>
<th>no reweighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRE</td>
<td>11.91</td>
<td>10.85</td>
<td>7.76</td>
<td>7.17</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0183</td>
<td>0.0208</td>
<td>0.0297</td>
<td>0.0317</td>
</tr>
</tbody>
</table>

Fig. 10. Singular values of the mode-$n$ unfoldings of a true third-order abundance tensor and its estimation by NL-TSUn for Example 1 with SNR = 20 dB. (a) Mode-1. (b) Mode-2. (c) Mode-3.

C. Convergence Analysis

Herein, we numerically show the convergence analysis when the ADMM is applied to solve the optimization problem in (13) for each group of similar HSI patches. We note that it is hard to guarantee the theoretical convergence analysis since the adoption of the reweighting strategies for $w_{l,n}$ and $z_{k,i}$ highlights the significance of simultaneous incorporation of the low-rank representation and the joint sparsity on the abundance tensors for hyperspectral unmixing problem.
Fig. 11. Horizontal slices of (top row) a true third-order abundance tensor and (bottom row) its estimation by NL-TSUn for Example 1 with SNR = 20 dB. The slice numbers (from left to right) are: 1, 13, 25, 313, 325, 601, 613, and 625.

Fig. 12. Objective function value versus iteration for Examples 1 and 2 when the ADMM is applied to solve (13).

Fig. 13. Plot of SRE (dB) against iteration of NL-TSUn for Example 1 with SNR = 20 dB.

$z_{k,i}$ renders the model (13) nonconvex. Fig. 12 depicts the objective function value versus iteration for Examples 1 and 2. Here for each noise level, the plot of one group of similar HSI patches is presented. Other groups show similar behavior so we omit them here. We see from Fig. 12 that, despite the fact that the convergence is not yet theoretically guaranteed, the convergence behavior becomes stable as the iteration number increases.

D. Parameter Discussion

In this section, we first test different maximum iteration numbers and then show the role of regularization parameters $\lambda_1$ and $\lambda_2$ in NL-TSUn. Fig. 13 plots the SRE value against iteration for Example 1 with SNR = 20 dB. We see from Fig. 13 that setting the maximum iteration number to be 500 is enough to obtain satisfactory estimations. Fig. 14 plots SRE (dB) values as a function of parameters $\lambda_1$ and $\lambda_2$ for Example 1. We see from the figure that both optimal $\lambda_1$ and $\lambda_2$ decrease as SNR gets higher. Optimal values for

Fig. 14. SRE (dB) as a function of regularization parameters $\lambda_1$ and $\lambda_2$ in NL-TSUn for Example 1.

Fig. 15. USGS map showing the location of different minerals in the Cuprite mining district in Nevada.
Fig. 16. Qualitative comparison between the classification maps of the AVIRIS Cuprite subscene produced by Tetracorder 4.4 and the abundance maps estimated by different unmixing algorithms for (top row) Alunite and (bottom row) Chalcedony. From left to right: Tetracorder 4.4, SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, JSpBLRU, and NL-TSUn.

Fig. 17. Classification maps of the regions in red boxes from Fig. 16 for (the first two rows) Alunite and (the last two rows) Chalcedony. From left to right: Tetracorder 4.4, SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, JSpBLRU, and NL-TSUn.

\( \lambda_2 \) are greater than or equal to those for \( \lambda_1 \). We also see that optimal parameter values are positive for all SNR levels, saying again the effectiveness of both low-rank and joint-sparse-based regularizations in NL-TSUn for hyperspectral unmixing.

V. EXPERIMENT ON REAL DATA

In this example, we test different unmixing algorithms on the widely used Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite data set.\(^5\) We use a subscene containing 200 \( \times \) 170 pixels with 188 spectral bands. The spectral library \( A_3 \in \mathbb{R}^{188 \times 240} \) is generated from the USGS library which includes all exposed minerals of interest. Fig. 15 shows a mineral map produced in 1995 by USGS. Since the detailed ground truth information is unavailable, we just qualitatively assess obtained abundance maps by different unmixing algorithms, with reference to the Tetracorder 4.4 software product\(^6\) [56]. Similarly as in [11]–[13], [22], [28], the regularization parameters for all algorithms are set to 0.001 except for CLSUnSAL with \( \lambda = 0.01 \).

Fig. 16 shows estimated abundance maps by different unmixing algorithms and Tetracorder 4.4 software for Alunite and Chalcedony. Particularly, the regions in red boxes are zoomed in and displayed in Fig. 17. We see from the two figures that all unmixing algorithms delineate the regions considered as respective materials. Nevertheless, NL-TSUn provides fractional abundance maps with more clear backgrounds and captures more spatial details. Generally, by exploiting nonlocal spatial similarity of HSI and imposing the low-rankness constraint and the joint sparsity on fractional abundance tensors, NL-TSUn is effective for unmixing real hyperspectral data.

VI. CONCLUSION AND FUTURE WORK

In this article, we have proposed a nonlocal tensor-based unmixing algorithm. We propose to exploit nonlocal spatial information in the HSI, instead of that in abundance maps. In this vein, we first group similar patches in the HSI and then introduce a novel unmixing model for these groups. Particularly, we introduce the weighted tensor \( \ell_{2,1} \)-norm to describe

\(^6\) Available online: https://speclab.cr.usgs.gov/PAPERS/tetracorder/
the joint sparsity of third-order abundance tensors, i.e., the groups of similar abundances. Simultaneously, we exploit the low-rank property of abundance tensors through the weighted tensor trace norm. The reweighting strategies have been employed in the weighted norms to enhance the sparsity on singular values and the row sparsity in each horizontal slice. Simulated and real-data experiments have demonstrated the effectiveness of the proposed algorithm, compared with other related state-of-the-art unmixing techniques. The derivation of more computationally efficient nonlocal techniques is currently under investigation.

ACKNOWLEDGMENT

The authors would like to thank the editors and the three anonymous reviewers for their constructive, detailed, and helpful advice regarding this article.

REFERENCES


Jie Huang (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013. She is an Associate Professor with the School of Mathematical Sciences, UESTC. Her research interests include remote sensing and image processing.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xian Jiaotong University, Xi’an, China, in 1986, 1992, and 2001, respectively. He is a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications. Dr. Huang is an Editor of The Scientific World Journal, Advances in Numerical Analysis, the Journal of Applied Mathematics, the Journal of Pure and Applied Mathematics: Advances in Applied Mathematics, and the Journal of Electronic Science and Technology.

Xi-Le Zhao received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively. He is a Professor with the School of Mathematical Sciences, UESTC. His research interest mainly focuses on model-driven and data-driven methods for image processing problems. His homepage is https://zhaoxile.github.io/.

Liang-Jian Deng (Member, IEEE) received the B.S. and Ph.D. degrees in applied mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010 and 2016, respectively. From 2013 to 2014, he was a Joint-Training Ph.D. Student with Case Western Reserve University, Cleveland, OH, USA. In 2017, he was a Postdoc with Hong Kong Baptist University (HKBU), Hong Kong. In addition, he also stayed with the Isaac Newton Institute for Mathematical Sciences, Cambridge University, Cambridge, U.K., and HKBU for short visits. He is an Associate Professor with the School of Mathematical Sciences, UESTC. His research interests include the use of partial differential equations (PDE), optimization modeling, and deep learning to address several tasks in image processing, and computer vision, e.g., resolution enhancement and restoration.