

Row-Sparsity Spectral Unmixing via Total Variation

Jin-Ju Wang , Ting-Zhu Huang , Jie Huang , Hong-Xia Dou, Liang-Jian Deng, and Xi-Le Zhao 

Abstract—Row sparsity of hyperspectral unmixing has attracted considerable attention in recent years. It exploits the fact that a hyperspectral image contains only a few number of endmembers. This property has been well investigated by the $\ell_{2,0}$ norm-based algorithms. The $\ell_{2,0}$ norm, however, is sensitive to noise in some cases. Therefore, we propose an unmixing model, which contains a total variation (TV) regularization and the $\ell_{2,0}$ norm, to address the drawback of the $\ell_{2,0}$ norm in the noise scenario, and promote piecewise smoothness in abundance maps. To solve the proposed model, we design an algorithm, termed as row-sparsity spectral unmixing via total variation (RSSUn-TV), under the nonconvex alternating direction method of multipliers (ADMM) framework. Particularly, we establish the convergence analysis of the RSSUn-TV algorithm. Experimental results on both synthetic and real hyperspectral data demonstrate that our proposed algorithm is effective for hyperspectral unmixing.

Index Terms—Nonconvex alternating direction method of multipliers (ADMM), $\ell_{2,0}$ regularization, row sparsity, spectral unmixing (SU), total variation (TV).

I. INTRODUCTION

HYPERSPECTRAL images are three-dimensional image data, which consists of hundreds of continuous-band images. It has been widely used in practical applications such as pharmaceutical process monitoring biomedical applications, and so on [1], [2]. However, due to the low spatial resolution of a sensor and the diversity of ground materials, each pixel in the hyperspectral image often contains more than one pure substance. These pixels, called mixed pixels, greatly affect the recognition and classification accuracy of ground materials. Therefore, spectral unmixing (SU), decomposing a mixed pixel into a collection of constituent spectra, called endmember, and their corresponding fractional abundances, called abundances, has become the key to the quantitative analysis of hyperspectral images.

SU algorithms primarily depend on the expected type of mixing, which can be illustrated as either a nonlinear mixing model or a linear mixing model (LMM) [1]. Among them, the LMM has been widely applied to the hyperspectral unmixing

Manuscript received July 28, 2019; revised October 2, 2019; accepted October 26, 2019. Date of publication November 18, 2019; date of current version February 4, 2020. This work was supported in part by NSFC under Grants 61772003, 61876203, 61702083, Science Strength Promotion Programme of UESTC and in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2019J093. (Corresponding author: Jin-Ju Wang.)

The authors are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: jinjuwang123@163.com; tingzhuhuang@126.com; happyjie07mo@163.com; 18483680787@163.com; liangjian1987112@126.com; xlzhao122003@163.com).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2950700

problem for simplicity and efficiency [3]–[6]. Specifically, the LMM assumes that the spectral response of a pixel is given by a linear combination of the endmembers present in the pixel and the corresponding fractional abundances [7]. We consider the LMM for the SU problem.

For the SU problem, notice that the LMM first identifies a collection of endmember, which is a difficult process. To overcome this drawback, a semisupervised strategy has been investigated in the literatures [3], [8], [9]. It replaces the collection of endmembers selected from images with known spectral libraries. This semisupervised strategy assumes that a mixed pixel only contains a few endmembers comparing with the large libraries. Thus, the fractional abundances of each pixel only have a few nonzeros. This property is modeled by the ℓ_1 norm in a sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) [3]. Furthermore, under the assumption of the semisupervised strategy, the number of endmembers in the hyperspectral image is much smaller than the number of endmembers in the library [8]. It leads to the abundance matrix to only have a few nonzero rows, which is called as row sparsity. Then, the collaborative SUnSAL (CLSUUnSAL) algorithm exploits the row sparsity by $\ell_{2,1}$ norm [8]. Recently, to describe the row sparsity directly, the $\ell_{2,0}$ norm, counting the number of nonzero rows in a matrix, is proposed in the collaborative sparse hyperspectral unmixing algorithm using ℓ_0 norm (CSUnL0) [10]. The CSUnL0 obtains an achievable unmixing results. Nevertheless, this method is hard to obtain accurate estimations in the high-level noisy environment due to the low noise tolerance of the $\ell_{2,0}$ norm [11].

In addition, to further improve the unmixing performance, many unmixing algorithms exploit spatial-contextual information under the sparse regression framework. Typically, the total variation (TV) regularization is one of the strategies for investigating the spatial information. The TV regularization was first proposed by Rudin *et al.* [12] for image denoising, which aims to describe the piecewise smooth property of natural images. It is a classic denoising regularization and has been verified for its effectiveness and stability in many denoising works [13]–[15]. For hyperspectral unmixing, TV is also an effective constraint of local spatial structure to make the abundance maps robust to noise and piecewise smooth [9], [15]–[17]. Typically, a sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV) is developed, which depicts the spatial consistency of the abundance maps by the TV regularization [9]. In [15], a TV-based regularization has also been utilized for modeling smoothness of abundance maps. In [16], the TV regularization is adopted to capture the piecewise smooth structure of each abundance map

and also regarded as an abundance map denoising procedure. All of them demonstrate the effectiveness of TV as a constraint in hyperspectral denoising and hyperspectral unmixing.

In this article, we propose an algorithm called row-sparsity spectral unmixing via total variation (RSSUn-TV). Remark that a pixel in a hyperspectral image consists of a small few end-members. We therefore use the $\ell_{2,0}$ regularization to promote the row sparsity in the abundance matrix. To capture the piecewise smooth structure of the abundance maps, we utilize the TV regularization as the constraint for the $\ell_{2,0}$ regularization-based sparse unmixing. Furthermore, the TV regularization is also regarded as an abundance map denoising procedure, which overcomes the drawbacks of the $\ell_{2,0}$ regularization in the noise scenario. Therefore, we propose an unmixing model, which contains the TV regularization and the $\ell_{2,0}$ regularization. To solve the proposed model, we design an algorithm, termed as RSSUn-TV, under the nonconvex alternating direction method of multipliers (ADMM) framework. Particularly, we present some analysis about the convergence of the RSSUn-TV algorithm. In addition, we employ three synthetic and one real-data hyperspectral images to evaluate the performance of the RSSUn-TV algorithm.

The structure of this article is outlined as follows. Section II introduces some sparse unmixing models. In Section III, we propose the given model and design a new unmixing algorithm to solve it. Section IV presents some theorems which support the proposed method in this article. The effectiveness of the proposed algorithm is demonstrated by both simulated experiments in Section V and a real data test in Section VI. Finally, Section VII concludes this article.

II. SPARSE UNMIXING MODEL

Assuming that the hyperspectral image $\mathbf{Y} = [y_1, y_2, \dots, y_n] \in \mathbb{R}^{L \times n}$ contains L spectral bands and n pixels for each band, y_i ($i = 1, 2, \dots, n$) is the i th observed spectral vector of a pixel in the hyperspectral image. The sparse unmixing model according to LMM is written as follows:

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N} \quad (1)$$

where $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$, containing m endmembers, is the abundance matrix, $\mathbf{N} \in \mathbb{R}^{L \times n}$ is the system noise, and x_i ($i = 1, 2, \dots, n$) is the i th abundance vector. According to the physical meaning of abundances, the abundance matrix often satisfies *abundance nonnegativity constraint* (ANC, i.e., $\mathbf{X} \geq \mathbf{0}$) and the *abundance sum-to-one constraint* (ASC, i.e., $\mathbf{1}_m^T \mathbf{X} = \mathbf{1}_n^T$) [18]. However, the ASC constraint is often not applicable for sparse unmixing for some reasons [3]. For example, the spectral library \mathbf{A} selection has certain randomness; it does not necessarily cover all the endmember lines in the target hyperspectral image. In practice, the number of endmembers in spectral library \mathbf{A} is much larger than the number of endmembers in a mixed pixel; the abundance matrix \mathbf{X} is sparse. Then, a sparse unmixing model based on sparsity is expressed as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X}\|_0 \text{ s.t. } \mathbf{X} \geq \mathbf{0} \quad (2)$$

where $\|\cdot\|_F$ represents the Frobenius norm, $\lambda \geq 0$ is the regularization parameter, and $\|\mathbf{X}\|_0$ represents the ℓ_0 quasi-norm of \mathbf{X} . According to the theory of compressed sensing and sparse representation, the ℓ_0 norm better depicts the sparsity, but the ℓ_0 minimization problem is nonconvex, and is difficult to solve [19]. From [20], [21], we see that the ℓ_0 norm optimization problem has the same solution as the ℓ_1 norm optimization problem under a certain condition of the restricted isometric property. And the ℓ_1 norm optimization problem is convex, making the solution is easier to get. Besides, using a weighted ℓ_1 regularizer to replace the ℓ_0 norm can provide better accuracy in some cases [11], [22]. In this context, the previous problem is written as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{W} \odot \mathbf{X}\|_1 \text{ s.t. } \mathbf{X} \geq \mathbf{0} \quad (3)$$

where $\|\mathbf{X}\|_1 = \sum_{j=1}^n \|x_j\|_1$, and x_j represents the j th column of the abundance matrix \mathbf{X} , \odot means the element-wise multiplication (i.e., Hadamard product) of two variables.

- 1) When $\mathbf{W} = \mathbf{I}$, then the model in (3) becomes the SUNSAL model in [3].
- 2) When \mathbf{W} is defined as

$$\mathbf{W}^{(t+1)} = \frac{1}{|\mathbf{X}|^{(t)} + \sigma}$$

where we call σ as a stabilization parameter. It aims at enhancing the sparsity of fractional abundances of mixed pixels.

In practice, a hyperspectral image always contains a small number of endmembers included in \mathbf{A} [8]. That means that the abundance matrix \mathbf{X} with the number of columns equal to the number of pixels only contains a few lines with nonzero entries. To implement this prior, collaborative sparse unmixing models are written as follows:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X}\|_{2,p} \text{ s.t. } \mathbf{X} \geq \mathbf{0}. \quad (4)$$

In the following, with different values of p , we obtain different unmixing models:

- 1) Let $p = 1$, then the model in (4) becomes the CLSUnSAL model in [8], where $(\|\mathbf{X}\|_{2,1} = \sum_{k=1}^m \|x^k\|_2)$, and x^k represents the k th row of \mathbf{X} , for $k = 1, \dots, m$. It promotes the sparsity of \mathbf{X} along the line.
- 2) Let $p = 0$, then the model in (4) becomes the CSUnL0 model in [10], where $\|\mathbf{X}\|_{2,0}$ counts the number of nonzero rows in matrix \mathbf{X} , $(\|\mathbf{X}\|_{2,0} = \sum_{k=1}^m \mathbf{1}(\|x^k\|_2 > 0))$, and x^k represents the k th row of \mathbf{X} , for $k = 1, \dots, m$. It promotes the sparsity of \mathbf{X} along the row by solving the sparse unmixing problem directly without replacing ℓ_0 norm by ℓ_1 norm.

III. PROPOSED MODEL AND ALGORITHM

In practice, the abundance matrix \mathbf{X} shares the same active set of endmembers [8]. To implement this prior, $\ell_{2,0}$ norm is used instead of $\ell_{2,1}$ norm. It counts the number of nonzero rows in matrix \mathbf{X} to describe the collaborative sparsity of the abundance matrix. However, $\ell_{2,0}$ has a low tolerance for noise [11]. Recall

that Rudin and Osher [12] proposed the TV model for achieving in removing image noise and keep details of image features, and it is known that the TV term as a spatial regularization is effective for hyperspectral unmixing. Therefore, in this article, we combine the $\ell_{2,0}$ regularization with the TV term to simultaneously exploit spatial information in hyperspectral images and suppress the noise, obtaining the following model:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X}\|_{2,0} + \lambda_{TV} \mathbf{TV}(\mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X} \geq \mathbf{0} \end{aligned} \quad (5)$$

where $\lambda_{TV} \geq 0$ is regularization parameter and $\mathbf{TV}(\mathbf{X}) = \|\mathbf{H}\mathbf{X}\|_1$ is a vector extension of the nonisotropic TV [9]. Here, $\mathbf{H} = [\mathbf{H}_h; \mathbf{H}_v]$, $\mathbf{H}_h : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ and $\mathbf{H}_v : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$, are horizontal and vertical differential operators, respectively. The $\mathbf{H}_h \mathbf{X}$ computes the differences between the components of \mathbf{X} and the corresponding right-side adjacent pixels with cyclic boundary assumption, and the same way for $\mathbf{H}_v \mathbf{X}$, which corresponds to the differences with the up-side adjacent pixels.

Recall that (5) is nonconvex and the nonconvex ADMM is introduced to solve nonconvex and nonsmooth optimization problem in [23]. In addition, Wang *et al.* provide sufficient conditions for the nonconvex ADMM algorithm to converge. Now, we will show in detail how to solve (5) by nonconvex ADMM.

To this end, we first equivalently rewrite (5) by some substitutions, see as follows:

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{V}_5} \quad & \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{V}_2\|_{2,0} \\ & + \lambda_{TV} \|\mathbf{V}_4\|_{1,1} + l_{R+}(\mathbf{V}_5) \\ \text{s.t.} \quad & \mathbf{AX} = \mathbf{V}_1, \quad \mathbf{X} = \mathbf{V}_2, \quad \mathbf{X} = \mathbf{V}_3 \\ & \mathbf{HV}_3 = \mathbf{V}_4, \quad \mathbf{X} = \mathbf{V}_5 \end{aligned} \quad (6)$$

where $l_{R+}(\mathbf{X}) = \sum_{i=1}^n l_{R+}(x_i)$ represents the nonnegative constraint of $\mathbf{X} \geq \mathbf{0}$. For $x_i \geq 0$, $l_{R+}(x_i) = 0$, and otherwise, $l_{R+}(x_i) = \infty$. Define

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \\ \mathbf{V}_4 \\ \mathbf{V}_5 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{A} \\ \mathbf{I} \\ \mathbf{I} \\ \mathbf{I} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -\mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & -\mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & -\mathbf{I} & 0 & 0 \\ 0 & 0 & \mathbf{H} & -\mathbf{I} & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{I} \end{pmatrix}$$

and

$$\begin{aligned} g(\mathbf{V}) = \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{V}_2\|_{2,0} \\ + \lambda_{TV} \|\mathbf{V}_4\|_{1,1} + l_{R+}(\mathbf{V}_5). \end{aligned}$$

Then, we reformulate (6) to a compact form

$$\min_{\mathbf{X}, \mathbf{V}} g(\mathbf{V}) \quad \text{s.t.} \quad \mathbf{GX} + \mathbf{BV} = \mathbf{0}.$$

To solve the above model by the nonconvex ADMM, we first introduce Lagrange multipliers

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \mathbf{D}_3 \\ \mathbf{D}_4 \\ \mathbf{D}_5 \end{pmatrix}$$

and define the augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{X}, \mathbf{V}, \mathbf{D}) = g(\mathbf{V}) - \frac{\mu}{2} \sum_{i=1}^5 \|\mathbf{D}_i\|_F^2 \\ + \frac{\mu}{2} \|\mathbf{GX} + \mathbf{BV} - \mathbf{D}\|_F^2 \end{aligned} \quad (7)$$

where $\mu > 0$ is a penalty parameter. Then, we minimize $\mathcal{L}_\mu(\mathbf{X}, \mathbf{V}, \mathbf{D})$ with respect to \mathbf{X} and \mathbf{V} , and update \mathbf{D} as the following framework:

$$\begin{cases} \mathbf{X}^{(t+1)} = \arg \min_{\mathbf{X}} \mathcal{L}_\mu(\mathbf{X}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) \\ \mathbf{V}^{(t+1)} = \arg \min_{\mathbf{V}} \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}, \mathbf{D}^{(t)}) \\ \mathbf{D}^{(t+1)} = \mathbf{D}^{(t)} - \mathbf{GX}^{(t+1)} - \mathbf{BV}^{(t+1)}. \end{cases}$$

In the following, we compute \mathbf{X}, \mathbf{V} -problems by the above-mentioned framework. To begin, for the \mathbf{X} subproblem, we have

$$\begin{aligned} \mathbf{X}^{(t+1)} = \arg \min_{\mathbf{X}} \frac{\mu}{2} \|\mathbf{AX} - \mathbf{V}_1^{(t)} - \mathbf{D}_1^{(t)}\|_F^2 \\ + \sum_{i=2,3,5} \frac{\mu}{2} \|\mathbf{X} - \mathbf{V}_i^{(t)} - \mathbf{D}_i^{(t)}\|_F^2. \end{aligned}$$

It is easy to obtain that

$$\begin{aligned} \mathbf{X}^{(t+1)} = (\mathbf{A}^T \mathbf{A} + 3\mathbf{I})^{-1} \left(\mathbf{A}^T \left(\mathbf{V}_1^{(t)} + \mathbf{D}_1^{(t)} \right) \right. \\ \left. + \sum_{i=2,3,5} \left(\mathbf{V}_i^{(t)} + \mathbf{D}_i^{(t)} \right) \right). \end{aligned}$$

Next, we decouple the \mathbf{V} subproblem into five parts. To compute \mathbf{V}_1 , we have the optimization problem

$$\begin{aligned} \mathbf{V}_1^{(t+1)} \\ = \arg \min_{\mathbf{V}_1} \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \frac{\mu}{2} \|\mathbf{AX}^{(t+1)} - \mathbf{V}_1 - \mathbf{D}_1^{(t)}\|_F^2. \end{aligned}$$

A simple calculation gives

$$\mathbf{V}_1^{(t+1)} = \frac{1}{1 + \mu} \left(\mathbf{Y} + \mu \left(\mathbf{AX}^{(t+1)} - \mathbf{D}_1^{(t)} \right) \right).$$

Then, \mathbf{V}_2 is computed by solving the optimization problem

$$\mathbf{V}_2^{(t+1)} = \arg \min_{\mathbf{V}_2} \lambda \|\mathbf{V}_2\|_{2,0} + \frac{\mu}{2} \|\mathbf{X}^{(t+1)} - \mathbf{V}_2 - \mathbf{D}_2^{(t)}\|_F^2.$$

Before studying the \mathbf{V}_2 subproblem, we first introduce a row-hard-threshold function [10], which is defined as follows:

$$(\mathbf{T}_{row}^{VHT}(\mathbf{B}, w))_{(i,:)} = \begin{cases} \mathbf{0}, & \|\mathbf{B}(i,:)\|_2 \leq w \\ \mathbf{B}(i,:), & \|\mathbf{B}(i,:)\|_2 > w \end{cases} \quad (8)$$

where w is a nonnegative parameter, and $\mathbf{B}(i,:)$ represents the i th row of \mathbf{B} for any matrix $\mathbf{B} \in \mathbb{R}^{m \times n}$. Thus, we get

$$\mathbf{V}_2^{(t+1)} = \mathbf{T}_{row}^{VHT} \left(\mathbf{X}^{(t+1)} - \mathbf{D}_2^{(t)}, \sqrt{\frac{2\lambda}{\mu}} \right).$$

In order to compute \mathbf{V}_3 , we have

$$\begin{aligned} \mathbf{V}_3^{(t+1)} &= \arg \min_{\mathbf{V}_3} \frac{\mu}{2} \|\mathbf{X}^{(t+1)} - \mathbf{V}_3 - \mathbf{D}_3^{(t)}\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{HV}_3 - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t)}\|_F^2. \end{aligned}$$

It is easy to obtain that

$$\begin{aligned} \mathbf{V}_3^{(t+1)} &= (\mathbf{H}^T \mathbf{H} + \mathbf{I})^{-1} \left(\mathbf{H}^T \left(\mathbf{V}_4^{(t)} + \mathbf{D}_4^{(t)} \right) \right. \\ &\quad \left. + \mathbf{X}^{(t+1)} - \mathbf{D}_3^{(t)} \right). \end{aligned}$$

For the \mathbf{V}_4 subproblem, we have

$$\begin{aligned} \mathbf{V}_4^{(t+1)} &= \arg \min_{\mathbf{V}_4} \lambda_{TV} \|\mathbf{V}_4\|_{1,1} + \frac{\mu}{2} \|\mathbf{HV}_3^{(t+1)} \\ &\quad - \mathbf{V}_4 - \mathbf{D}_4^{(t)}\|_F^2. \end{aligned}$$

Then it is easy to get

$$\mathbf{V}_4^{(t+1)} = \text{soft} \left(\mathbf{HV}_3^{(t+1)} - \mathbf{D}_4^{(t)}, \frac{\lambda_{TV}}{\mu} \right)$$

where $\text{soft}(u, \tau)$, defined by

$$\text{soft}(u, \tau) = \text{sign}(u) \max\{|u| - \tau, 0\}$$

is a component-wise application of the *soft-threshold* function in [24], [25].

Finally, for \mathbf{V}_5 , we have

$$\mathbf{V}_5^{(t+1)} = \arg \min_{\mathbf{V}_5} \iota_{R_+}(\mathbf{V}_5) + \frac{\mu}{2} \|\mathbf{X}^{(t+1)} - \mathbf{V}_5 - \mathbf{D}_5^{(t)}\|_F^2$$

and the solution is

$$\mathbf{V}_5^{(t+1)} = \max(\mathbf{X}^{(t+1)} - \mathbf{D}_5^{(t)}, \mathbf{0}).$$

Finally, we update all Lagrange multipliers as follows

$$\begin{cases} \mathbf{D}_1^{(t+1)} = \mathbf{D}_1^{(t)} - (\mathbf{AX}^{(t+1)} - \mathbf{V}_1^{(t+1)}) \\ \mathbf{D}_2^{(t+1)} = \mathbf{D}_2^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_2^{(t+1)}) \\ \mathbf{D}_3^{(t+1)} = \mathbf{D}_3^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_3^{(t+1)}) \\ \mathbf{D}_4^{(t+1)} = \mathbf{D}_4^{(t)} - (\mathbf{HV}_3^{(t+1)} - \mathbf{V}_4^{(t+1)}) \\ \mathbf{D}_5^{(t+1)} = \mathbf{D}_5^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_5^{(t+1)}) \end{cases}.$$

To make it more clearly, we summarize the proposed algorithm, named as RSSUn-TV algorithm, in Algorithm 1.

IV. CONVERGENCE ANALYSIS

Due to the nonconvexity of RSSUn-TV, we will discuss the convergence condition of our algorithm in Theorem 1. Before

Algorithm 1: The RSSUn-TV Algorithm.

- 01) **Initialization:** $\mathbf{X}^{(0)}, \mathbf{V}_1^{(0)} = \mathbf{AX}^{(0)}, \mathbf{V}_4^{(0)} = \mathbf{HV}_3^{(0)}, \mathbf{V}_i^{(0)} = \mathbf{X}^{(0)} (i = 2, 3, 5), \mathbf{D}^{(0)}, \varepsilon > 0, t=0.$
 - 02) Set $\lambda, \lambda_{TV}, \mu > 0$.
 - Repeat:**
 - 03) $\mathbf{X}^{(t+1)} = (\mathbf{A}^T \mathbf{A} + 3\mathbf{I})^{-1} (\mathbf{A}^T (\mathbf{V}_1^{(t)} + \mathbf{D}_1^{(t)}) + \sum_{i=2,3,5} (\mathbf{V}_i^{(t)} + \mathbf{D}_i^{(t)}))$
 - 04) $\mathbf{V}_1^{(t+1)} = \frac{1}{1+\mu} (\mathbf{Y} + \mu (\mathbf{AX}^{(t+1)} - \mathbf{D}_1^{(t)}))$
 - 05) $\mathbf{V}_2^{(t+1)} = \mathbf{T}_{row}^{VHT} (\mathbf{X}^{(t+1)} - \mathbf{D}_2^{(t)}, \sqrt{\frac{2\lambda}{\mu}})$
 - 06) $\mathbf{V}_3^{(t+1)} = (\mathbf{H}^T \mathbf{H} + \mathbf{I})^{-1} (\mathbf{A}^T (\mathbf{V}_4^{(t)} + \mathbf{D}_4^{(t)}) + (\mathbf{X}^{(t+1)} - \mathbf{D}_3^{(t)}))$
 - 07) $\mathbf{V}_4^{(t+1)} = \text{soft}(\mathbf{HV}_3^{(t+1)} - \mathbf{D}_4^{(t)}, \frac{\lambda_{TV}}{\mu})$
 - 08) $\mathbf{V}_5^{(t+1)} = \max(\mathbf{X}^{(t+1)} - \mathbf{D}_5^{(t)}, \mathbf{0})$
 - Update the Lagrange multipliers:**
 - 09) $\mathbf{D}_1^{(t+1)} = \mathbf{D}_1^{(t)} - (\mathbf{AX}^{(t+1)} - \mathbf{V}_1^{(t+1)})$
 - 10) $\mathbf{D}_2^{(t+1)} = \mathbf{D}_2^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_2^{(t+1)})$
 - 11) $\mathbf{D}_3^{(t+1)} = \mathbf{D}_3^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_3^{(t+1)})$
 - 12) $\mathbf{D}_4^{(t+1)} = \mathbf{D}_4^{(t)} - (\mathbf{HV}_3^{(t+1)} - \mathbf{V}_4^{(t+1)})$
 - 13) $\mathbf{D}_5^{(t+1)} = \mathbf{D}_5^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_5^{(t+1)})$
 - Update Number of iterations:**
 - 14) $t = t + 1$
 - Until termination condition** $\|\mathbf{GX}^{(t)} + \mathbf{BV}^{(t)}\|_F \leq \varepsilon$ **or the maximum iteration is met.**
-

giving Theorem 1, following are some definitions that will be used later.

Definition 1.(coercivity[23]): Define the feasible set $F := (\mathbf{X}, \mathbf{V}) : \mathbf{GX} + \mathbf{BV} = \mathbf{0}$. A function $T(\mathbf{X}, \mathbf{V})$ is coercive over this set, that is, $T(\mathbf{X}, \mathbf{V}) \rightarrow \infty$ if $(\mathbf{X}, \mathbf{V}) \in F$ and $\|(\mathbf{X}, \mathbf{V})\| \rightarrow \infty$.

Before introducing the convergence analysis of our algorithm, we first give the following assumption similarly as in [10].

Assumption 1: Assume that the abundance matrix always satisfies abundance non-negativity constraint (ANC, i.e., $\mathbf{X} \geq \mathbf{0}$).

With Assumption 1, we rewrite the original problem as follows:

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4} \quad & \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{V}_2\|_{2,0} + \lambda_{TV} \|\mathbf{V}_4\|_{1,1} \\ \text{s.t.} \quad & \mathbf{AX} = \mathbf{V}_1, \quad \mathbf{X} = \mathbf{V}_2, \quad \mathbf{X} = \mathbf{V}_3, \quad \mathbf{HV}_3 = \mathbf{V}_4. \end{aligned} \quad (9)$$

Then, we define

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \\ \mathbf{V}_4 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{A} \\ \mathbf{I} \\ \mathbf{I} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -\mathbf{I} & 0 & 0 & 0 \\ 0 & -\mathbf{I} & 0 & 0 \\ 0 & 0 & -\mathbf{I} & 0 \\ 0 & 0 & \mathbf{H} & -\mathbf{I} \end{pmatrix}$$

and

$$g(\mathbf{V}) = \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{V}_2\|_{2,0} + \lambda_{TV} \|\mathbf{V}_4\|_{1,1}.$$

The augmented Lagrange function of (9) is

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{X}, \mathbf{V}, \mathbf{D}) &= g(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{AX} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2 \\ &\quad - \frac{\mu}{2} \sum_{i=1}^4 \|\mathbf{D}_i\|_F^2 \\ &\quad + \sum_{i=2,3} \frac{\mu}{2} \|\mathbf{X} - \mathbf{V}_i - \mathbf{D}_i\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{HV}_3 - \mathbf{V}_4 - \mathbf{D}_4\|_F^2. \end{aligned} \quad (10)$$

Furthermore, the corresponding optimization for (10) is

$$\begin{aligned} \mathbf{X}^{(t+1)} &= (\mathbf{A}^T \mathbf{A} + 2\mathbf{I})^{-1} \\ &\quad \times \left(\mathbf{A}^T (\mathbf{V}_1^{(t)} + \mathbf{D}_1^{(t)}) + \sum_{i=2,3} (\mathbf{V}_i^{(t)} + \mathbf{D}_i^{(t)}) \right) \\ \mathbf{V}_1^{(t+1)} &= \frac{1}{1+\mu} \left(\mathbf{Y} + \mu (\mathbf{AX}^{(t+1)} - \mathbf{D}_1^{(t)}) \right) \\ \mathbf{V}_2^{(t+1)} &= \mathbf{T}_{row}^{VHT} \left(\mathbf{X}^{(t+1)} - \mathbf{D}_2^{(t)}, \sqrt{\frac{2\lambda}{\mu}} \right) \\ \mathbf{V}_3^{(t+1)} &= (\mathbf{H}^T \mathbf{H} + \mathbf{I})^{-1} \left(\mathbf{A}^T (\mathbf{V}_4^{(t)} + \mathbf{D}_4^{(t)}) \right. \\ &\quad \left. + (\mathbf{X}^{(t+1)} - \mathbf{D}_3^{(t)}) \right) \\ \mathbf{V}_4^{(t+1)} &= \text{soft} \left(\mathbf{HV}_3^{(t+1)} - \mathbf{D}_4^{(t)}, \frac{\lambda_{TV}}{\mu} \right). \end{aligned} \quad (11)$$

Theorem 1: Sequence $\{\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}\}$ generated by (11) is bounded and the augmented Lagrangian \mathcal{L}_μ in (10) is decreasing with a lower bound if

$$8m\lambda + 2(1 + \|\mathbf{H}\|_F^2) \frac{\lambda_{TV}^2}{\mu} \leq \frac{\mu^3}{2} \left(\varepsilon - 2 \left(\frac{\lambda_{TV}}{\mu} \right)^2 - \frac{4\lambda}{\mu} \right)^2$$

holds.

Proof: By the definition of coercivity, $g(\mathbf{V}) = \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{V}_2\|_{2,0} + \lambda_{TV} \|\mathbf{V}_4\|_{1,1}$ is coercive the set F . Since $(\mathbf{X}, \mathbf{V}) \in F$ and $\|(\mathbf{X}, \mathbf{V})\| \rightarrow \infty$, then $g(\mathbf{V}) \rightarrow \infty$. In Appendix A, we have proved \mathcal{L}_μ in (10) is decreasing. Now, the lower bound of the augmented Lagrange function is found as follows:

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{X}, \mathbf{V}, \mathbf{D}) &= g(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{AX} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2 - \frac{\mu}{2} \sum_{i=1}^4 \|\mathbf{D}_i\|_F^2 \\ &\quad + \sum_{i=2,3} \frac{\mu}{2} \|\mathbf{X} - \mathbf{V}_i - \mathbf{D}_i\|_F^2 + \frac{\mu}{2} \|\mathbf{HV}_3 - \mathbf{V}_4 - \mathbf{D}_4\|_F^2 \\ &\geq -\frac{\mu}{2} \sum_{i=1}^5 \|\mathbf{D}_i\|_F^2. \end{aligned} \quad (12)$$

From Appendix A, \mathcal{L}_μ is upper bounded by $\mathcal{L}_\mu(\mathbf{X}^0, \mathbf{V}^0, \mathbf{D}^0)$. By the coercivity of $\mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)})$, $\mathbf{V}^{(t)}$ is bounded

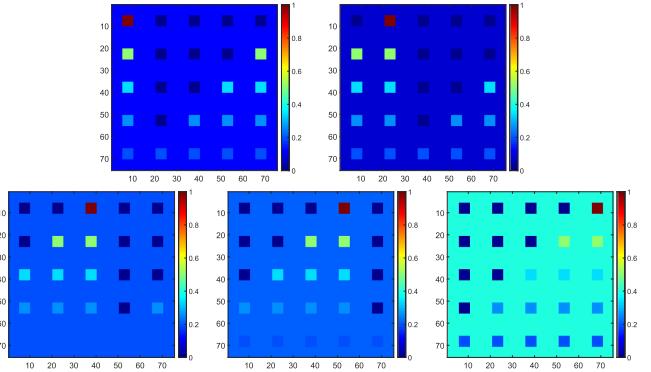


Fig. 1. True fractional abundances of five endmembers for DC1. The first row: endmembers #1-#2. The second row: endmembers #3-#5.

and, therefore, $\mathbf{X}^{(t)}$ and $\mathbf{D}^{(t)}$ are also bounded. Therefore, $\mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)})$ has a lower bounded as $t \rightarrow \infty$ ■

Based on Theorem 1, we know that the sequence $\{\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}\}$ is bounded. By Bolzano–Weierstrass theorem, the sequence has at least one accumulation point. We also remark that this theorem cannot guarantee that this accumulation point is a stationary point of problem (9), because the duality condition is not always true when the objective function is nonconvex.

V. SYNTHETIC DATA EXPERIMENTS

In this section, we perform unmixing experiments on two synthetic hyperspectral data cubes. We compare our method with four state-of-the-art unmixing algorithms: SUNSAL [3], CLSUNSAL [8], CLSUNL0 [10], and SUNSAL-TV [9]. All tests are implemented on the platform of Windows 7 and MATLAB (R2016a) with an Intel Core i5-4590, 3.30 GHz and 8 GB RAM.

A. Generation of Synthetic Data

We will construct three synthetic hyperspectral data cubes to carry on our experiments. For the simulated data cube 1 (DC1), we first choose the spectral library $\mathbf{A}_1 \in \mathbb{R}^{224 \times 240}$, which is generated by selecting 240 materials (different mineral types) from the United States Geological Survey (USGS) library¹ randomly. \mathbf{A}_1 includes spectral signatures with reflectance values given in 224 spectral bands, distributed uniformly in the interval $0.4 - 2.5 \mu\text{m}$. Then, we select five spectral signatures randomly from the library \mathbf{A}_1 as active endmembers. The corresponding true fractional abundances for these endmembers are shown in Fig. 1. After the above procedures, we generate the observed matrix $\mathbf{DC1} \in \mathbb{R}^{224 \times 5625}$ according to LMM.

For the simulated data cube 2 (DC2), we first choose the spectral library matrix, $\mathbf{A}_2 \in \mathbb{R}^{100 \times 120}$, which is generated by using a library of 262 spectral signatures generally found on satellites, from the National Aeronautics and Space Administration Johnson Space Center (NASA JSC) [26], with 100 spectral bands. Then, we choose nine signatures randomly from \mathbf{A}_2 as active endmembers, and the corresponding true fractional

¹ Available online: <http://speclab.cr.usgs.gov/spectral.lib06>

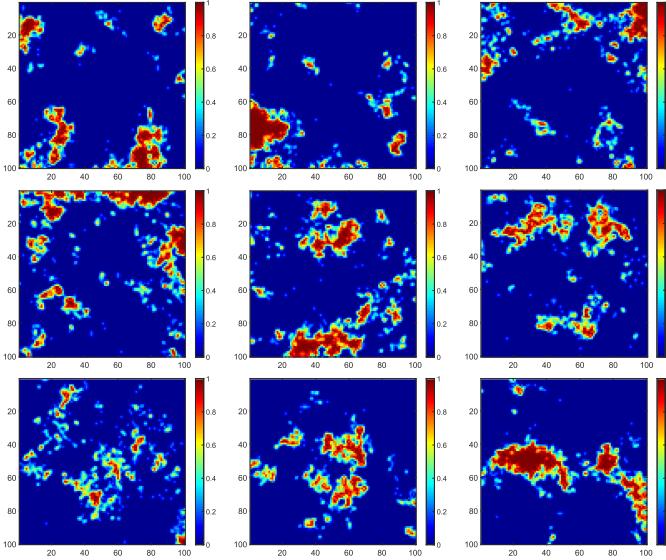


Fig. 2. True fractional abundances of nine endmembers for DC2. The first row: endmembers #1-#3. The second row: endmembers #4-#6. The third row: endmembers #7-#9.

abundances for each of the nine endmembers are shown in Fig. 2. Subsequently, we generate the observed matrix $\mathbf{DC2} \in \mathbb{R}^{100 \times 10000}$ by the LMM.

For the simulated data cube 3 (DC3), we consider a simulated hyperspectral image of the Hubble Space Telescope, similar to those used in [27], [28]. The data are a simulation of that collected by the U.S. Air Force AEOS Spectral Imaging Sensor (ASIS), which is used to collect adaptive optics compensated spectral images of astronomical objects and satellites at the Maui Space Surveillance Center for space object surveillance [29]. We select twelve signatures randomly from $\mathbf{A}_2 \in \mathbb{R}^{100 \times 120}$ as active endmembers. The corresponding true fractional abundances for these endmembers are shown in Fig. 3. Subsequently, we generate the observed matrix $\mathbf{DC3} \in \mathbb{R}^{100 \times 14400}$ according to LMM.

Finally, the above three data cubes are degraded by Gaussian noise with the signal-to-noise ratio (SNR), i.e.,

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{AX}\|_F^2}{\|\mathbf{N}\|_F^2}$$

of 30, 40, and 50 dB, respectively.

B. Experimental Settings

We adopt the signal to reconstruction error (SRE), measured in dB, as an index to evaluate the experimental results, similarly as in [3]. The SRE is defined as follows:

$$\text{SRE(dB)} = 10 \log_{10} \frac{\|\mathbf{X}_{true}\|_F^2}{\|\mathbf{X}_{true} - \hat{\mathbf{X}}\|_F^2}$$

where \mathbf{X}_{true} denotes the true fractional abundances, and $\hat{\mathbf{X}}$ is the computed fractional abundances by different algorithms. Generally, the higher the SRE (dB) value, the better unmixing performance of the algorithm is. We test the unmixing

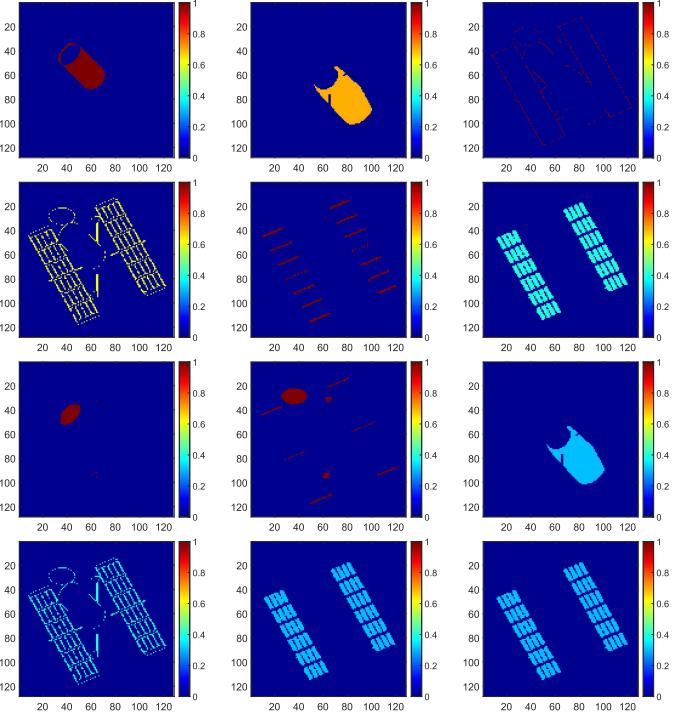


Fig. 3. True fractional abundances of nine endmembers for DC3. The first row: endmembers #1-#3. The second row: endmembers #4-#6. The third row: endmembers #7-#9. The fourth row: endmembers #10-#12.

algorithms: SUNSAL, CLSUnSAL, CLSUnL0, SUNSAL-TV, and our RSSUn-TV on the three data cubes using different values of λ from $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5\}$, λ_{TV} from $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$, and μ from $\{0.01, 0.05, 0.1, 0.5, 1, 2\}$. All possible combinations of these parameters are considered. For each experiment, we operate corresponding algorithms until

$$\|\mathbf{GX}^{(t)} + \mathbf{BV}^{(t)}\|_F \leq \varepsilon = 5 \times 10^{-6}$$

or the maximum iteration number reach at 1000 for a fair comparison. The initialization for RSSUn-TV is

$$\begin{aligned} \mathbf{X}^{(0)} &= (\mathbf{A}^T \mathbf{A} + 3\mathbf{I})^{-1} \mathbf{A}^T \mathbf{Y}, \quad \mathbf{V}_1^{(0)} = \mathbf{AX}^{(0)} \\ \mathbf{V}_4^{(0)} &= \mathbf{HV}_3^{(0)}, \quad \mathbf{V}_i^{(0)} = \mathbf{X}^{(0)} (i = 2, 3, 5) \end{aligned}$$

which is similarly as in SUNSAL-TV. In addition, the corresponding regularization parameters of each algorithm are tuned to their highest SRE (dB) values.

C. Experimental Results

In this section, we have done five experiments. The first two experiments are to analyze the noise sensitivity and the effectiveness of the $\ell_{2,0}$ norm, and other three experiments are to verify the effectiveness of the proposed algorithm for hyperspectral unmixing on the abovementioned three data cubes.

1) *Noise Sensitivity for the $\ell_{2,1}$ Norm and the $\ell_{2,0}$ Norm:* In this part, we study the noise tolerance of $\ell_{2,1}$ norm and $\ell_{2,0}$ norm. In this experiment, we compute the SRE(dB) values of the CLSUnSAL algorithm, the CSUnL0 algorithm, and our

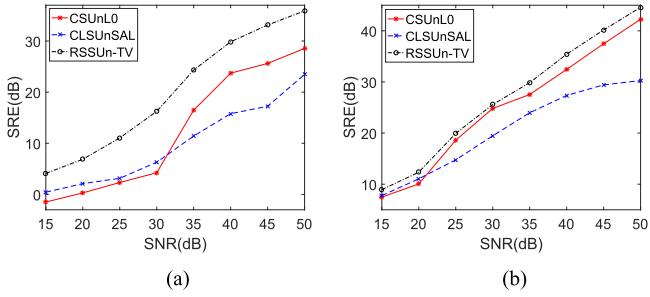


Fig. 4. SRE(dB) of CLSUnSAL, CSUnL0, and RSSUn-TV at different noise levels for (a) Gaussian noise and (b) colored noise on DC1.

TABLE I
SRE(dB) VALUES OF RSSUn-TV ON DC1 FOR DIFFERENT PARAMETER VALUES (OPTIMAL PARAMETERS FOR WHICH THE REPORTED VALUES WERE ACHIEVED ARE INDICATED)

SNR	RSSUn-TV with $\lambda_{TV} = 0$	RSSUn-TV with $\lambda = 0$ (only the $\ell_{2,0}$ prior)	RSSUn-TV with $\lambda = 0$ (only the TV prior)	RSSUn-TV
	4.37	9.68	16.24	
30dB	($\lambda = 0.5$)	($\lambda_{TV} = 0.01$)	($\lambda = 0.5$, $\lambda_{TV} = 0.01$)	
	21.24	19.24	29.84	
40dB	($\lambda = 0.5$)	($\lambda_{TV} = 0.005$)	($\lambda = 0.5$, $\lambda_{TV} = 0.005$)	
	26.46	20.70	35.89	
50dB	($\lambda = 0.05$)	($\lambda_{TV} = 0.005$)	($\lambda = 0.5$, $\lambda_{TV} = 0.001$)	

The bold entity means the highest SRE value at each case.

proposed algorithm at different Gaussian noise and colored noise levels on DC1 in Fig. 4. As shown in Fig. 4, the SRE(dB) values of CLSUnSAL are clearly better than those of CSUnL0 for both the high Gaussian and colored noise levels of SNR = 15 and 20 dB. That shows that the $\ell_{2,0}$ regularization is not good enough for the higher noise levels. Therefore, we conclude that $\ell_{2,0}$ norm has low tolerance for high noise level and is more sensitive to noise than $\ell_{2,1}$ norm for both Gaussian noise and colored noise. In addition, Fig. 4 shows that RSSUn-TV is beneficial for suppressing the noise, which is especially clear for Gaussian noise.

2) *Effectiveness of the $\ell_{2,0}$ Norm and the TV Term:* Here, we study the effectiveness of the $\ell_{2,0}$ norm and the TV term. In Table I, we show the SRE(dB) values of RSSUn-TV at different noise levels on DC1 for different parameter values. According to Table I, RSSUn-TV with $\lambda = 0$ obtains a higher SRE(dB) value than RSSUn-TV with $\lambda_{TV} = 0$ when SNR = 30 dB, while RSSUn-TV with $\lambda_{TV} = 0$ obtains higher SRE(dB) values for SNR = 40 and 50 dB. And RSSUn-TV always obtains higher SRE values than RSSUn-TV with $\lambda = 0$ and RSSUn-TV with $\lambda_{TV} = 0$ in any cases.

3) *Comparison With Other Unmixing Algorithms:* We now study the performance of the proposed RSSUn-TV and the other four compared algorithms with regard to different SNRs for the three data cubes. Table II shows the SREs obtained by different algorithms under different SNRs for three data cubes. In detail, the other five compared algorithms and the proposed RSSUn-TV are tuned to their best performances by adopting the

TABLE II
SRE(dB) VALUES BY DIFFERENT UNMIXING ALGORITHMS FOR DC1 AND DC2 (OPTIMAL PARAMETERS FOR WHICH THE REPORTED VALUES WERE ACHIEVED ARE INDICATED)

DC1					
SNR	SUnSAL	CLSUnSAL	CSUnL0	SUnSAL-TV	
30dB	6.12 ($\lambda = 0.5$)	6.30 ($\lambda = 0.05$)	4.24 ($\lambda = 0.5$)	13.68 ($\lambda = 0.01$, $\lambda_{TV} = 0.01$)	16.24 ($\lambda = 0.5$, $\lambda_{TV} = 0.01$)
40dB	11.04 ($\lambda = 0.01$)	15.79 ($\lambda = 0.5$)	23.71 ($\lambda = 0.5$)	21.44 ($\lambda = 0.01$, $\lambda_{TV} = 0.01$)	29.84 ($\lambda = 0.5$, $\lambda_{TV} = 0.005$)
50dB	20.71 ($\lambda = 0.005$)	23.47 ($\lambda = 0.05$)	26.45 ($\lambda = 1$)	27.88 ($\lambda = 0.001$, $\lambda_{TV} = 0.001$)	35.89 ($\lambda = 0.5$, $\lambda_{TV} = 0.001$)
DC2					
SNR	8.16 ($\lambda = 0.05$)	5.52 ($\lambda = 0.01$)	5.28 ($\lambda = 0.5$)	11.34 ($\lambda = 0.05$, $\lambda_{TV} = 0.001$)	14.01 ($\lambda = 0.5$, $\lambda_{TV} = 0.005$)
40dB	13.59 ($\lambda = 0.01$)	10.65 ($\lambda = 0.005$)	16.58 ($\lambda = 1$)	20.18 ($\lambda = 0.005$, $\lambda_{TV} = 0.001$)	25.86 ($\lambda = 0.5$, $\lambda_{TV} = 0.001$)
50dB	17.95 ($\lambda = 0.001$)	15.13 ($\lambda = 0.05$)	25.73 ($\lambda = 1$)	23.89 ($\lambda = 0.005$, $\lambda_{TV} = 0.001$)	29.44 ($\lambda = 0.5$, $\lambda_{TV} = 0.001$)
DC3					
SNR	5.19 ($\lambda = 0.1$)	8.27 ($\lambda = 0.1$)	9.02 ($\lambda = 0.5$)	15.12 ($\lambda = 0.01$, $\lambda_{TV} = 0.001$)	17.81 ($\lambda = 0.1$, $\lambda_{TV} = 0.005$)
40dB	16.86 ($\lambda = 0.01$)	18.34 ($\lambda = 0.01$)	19.58 ($\lambda = 0.5$)	21.92 ($\lambda = 0.005$, $\lambda_{TV} = 0.001$)	30.07 ($\lambda = 0.1$, $\lambda_{TV} = 0.001$)
50dB	27.59 ($\lambda = 0.005$)	22.87 ($\lambda = 0.01$)	37.92 ($\lambda = 0.01$)	39.24 ($\lambda = 0.005$, $\lambda_{TV} = 0.005$)	41.79 ($\lambda = 0.1$, $\lambda_{TV} = 0.001$)

The bold entity means the highest SRE value at each case.

parameter setting in Table II. According to Table II, the performances of all the algorithms degrade as the noise gets stronger. CSUnL0 provides lower SRE values than other five algorithms when SNR = 30 dB for DC1 and DC2, while its values are higher than the values obtained by SUnSAL and CLSUnSAL when SNR = 40, 50 dB. And CSUnL0 provides higher SRE values than the values obtained by SUnSAL and CLSUnSAL for DC3. Clearly, SUnSAL-TV and RSSUn-TV provide higher SRE values than the other three compared algorithms, and the RSSUn-TV outperforms other algorithms in terms of SRE at different noise levels.

To further verify the effectiveness of RSSUn-TV, we show the estimated abundance maps by five comparing algorithms for endmembers #3 and #5 on DC1, endmembers #4 and #7 on DC2, and endmembers #4 and #6 on DC3 with SNR = 30 dB in Figs. 5, 6, and 7, respectively. The conclusions are similar for other cases. Figs. 5, 6, and 7 demonstrate that the abundances estimated by SUnSAL, CLSUnSAL, and CSUnL0 exhibited similar behavior and their backgrounds look very noisy. Other two comparing algorithms, owing to the TV term, provide much smoother background. Though the abundances obtained by SUnSAL-TV, and RSSUn-TV appear very similar visually, the proposed RSSUn-TV retrieves more details (compare the reconstructed squares in the third and the fourth row of endmember #3 and in the second row of endmember #5 in Fig. 5). It is also clear from Fig. 6 that RSSUn-TV gives not only the highest abundances for the respective endmembers but also

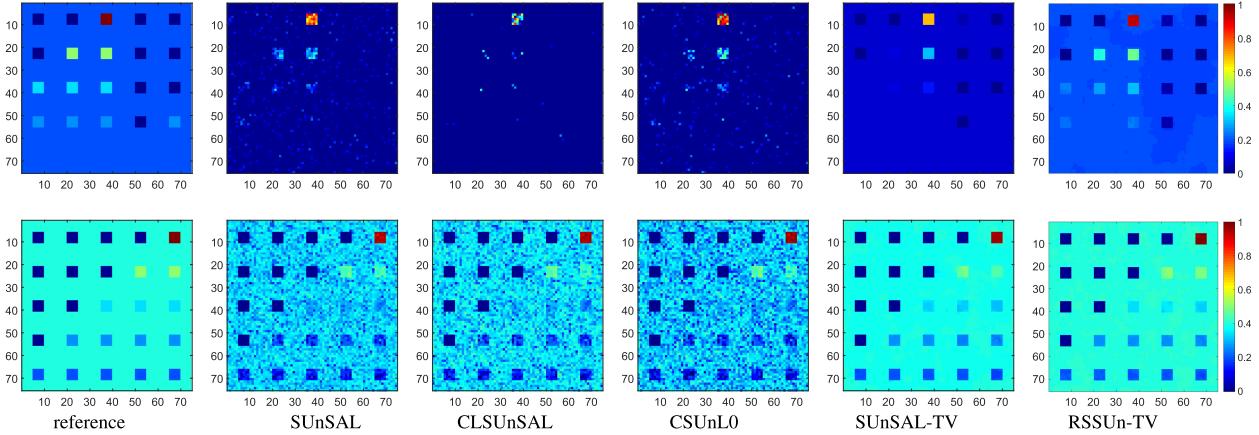


Fig. 5. Estimated abundance maps by different unmixing algorithms for (Top row) endmembers #3 and (Bottom row) #5 for DC1 with SNR = 30 dB.

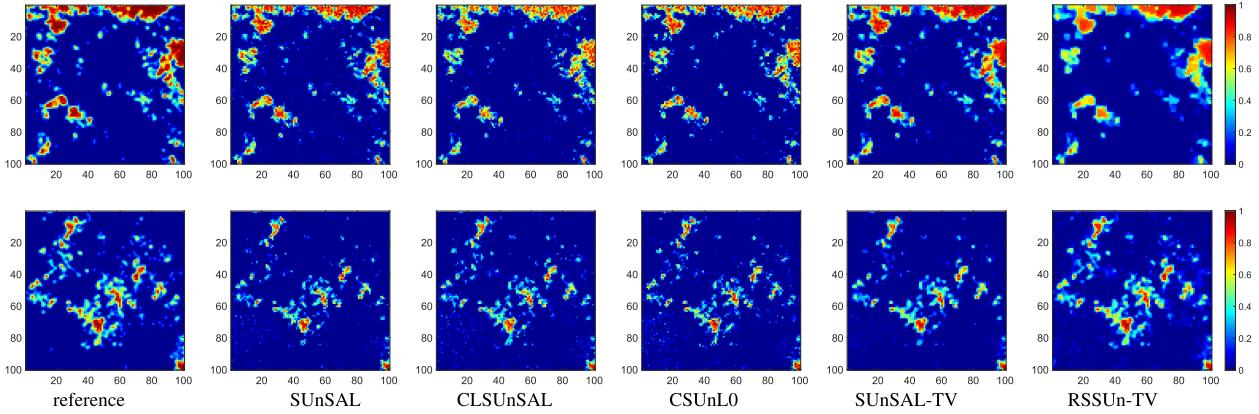


Fig. 6. Estimated abundance maps by different unmixing algorithms for (Top row) endmembers #4 and (Bottom row) #7 for DC2 with SNR = 30 dB.

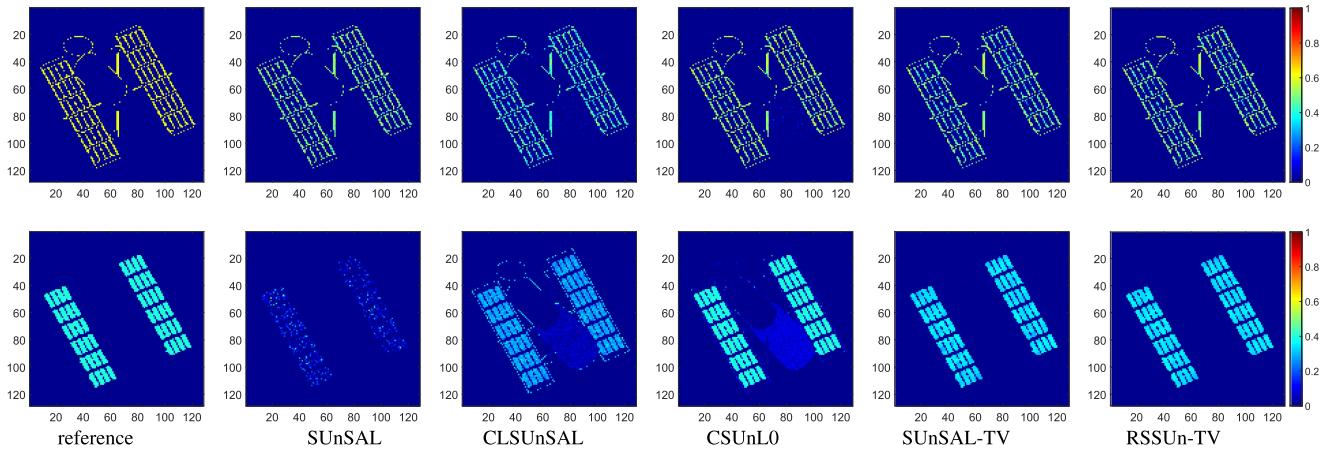


Fig. 7. Estimated abundance maps by different unmixing algorithms for (Top row) endmembers #4 and (Bottom row) #6 for DC3 with SNR = 30 dB.

the smoothest background. According to Fig. 7, RSSUn-TV and SUnSAL-TV provide similar abundance map corresponding to endmember #4, and RSSUn-TV is more similar with the true abundance map for endmember #6. In order to better visualize the effect of the RSSUn-TV, Fig. 8 shows true and estimated abundances by different algorithms for 100 selected pixels in

DC2 with SNR of 30 dB. From Fig. 8, although the first comparing three algorithms: SUnSAL, CLSUnSAL, and CLSUnL0 approximate the positions and values of the predicted line for 100 selected pixels. There are still many low abundance values that are not actually present in the image. On the contrary, the estimated abundances of other three algorithms are more

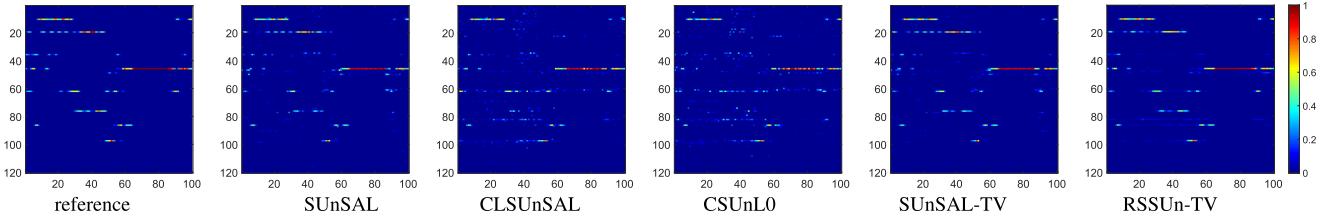


Fig. 8. True and estimated abundances by different algorithms for 100 selected pixels in DC2 with SNR of 30 dB. Vertical coordinate denotes endmembers, and horizontal coordinate denotes pixels.

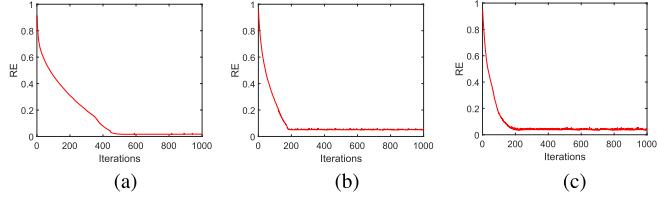


Fig. 9. Convergence curves of the proposed algorithm for: (a) DC1, (b) DC2, and (c) DC3 with SNR = 40 dB.

accurate and have a better visual effect. Clearly, RSSUn-TV provides the most similar abundances to the ground truth than other competing algorithms.

4) *Convergence Analysis*: The goal of this experiment is to numerically analyze the convergence of RSSUn-TV. To this end, Fig. 9 plots the relative error (RE) versus the iteration number for the three data cubes with SNR of 40 dB. The RE is defined as follows:

$$RE = \frac{\|\mathbf{X}_{true} - \hat{\mathbf{X}}\|_F^2}{\|\mathbf{X}_{true}\|_F^2}.$$

It is shown in Fig. 9 that the convergence curves practically tend to be stable.

5) *Parameter Discussion*: In this experiment, we study the selection regularization parameter of the proposed model. Fig. 10 shows SRE (dB) when these two parameters λ and λ_{TV} change for three data cubes with different SNR levels. Fig. 10 indicates that these two parameters λ and λ_{TV} have a great effect on the unmixing performance. For DC1 and DC2, with the increase of λ , the value of SRE increases. And the optimal λ is stable with three different SNR levels for the same data cube. For these three data cubes, the optimal λ_{TV} is fixed at the set {0.01, 0.001, 0.005}. Generally, the optimal λ is 0.5 or 0.1, and the optimal λ_{TV} is 0.001 or 0.005.

In order to analyze the change of optimal parameters: λ and λ_{TV} for different data, we generate 40 hyperspectral data cubes. We generate \mathbf{AX} by using DC1, and use the matlab function “randn” to generate 40 different \mathbf{N} s with the SNR = 30 dB. Then, we test the values of λ and λ_{TV} corresponding to the optimal SREs for these 40 data cubes, and show the change of λ , λ_{TV} , and SRE, respectively, in Fig. 11. According to Fig. 11, the optimal λ is 0.1 or 0.5, which is consistent with the result in Table II. And the optimal λ_{TV} is stable for different data cubes. The mean and variance of SRE are 16.87 and 0.4173, respectively. Generally, the SREs are stable for 40 different data cubes. Considering the uncertainty of parameters in different

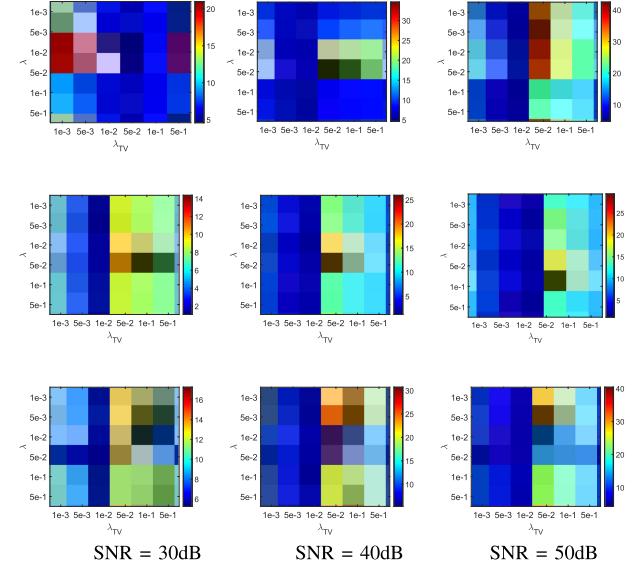


Fig. 10. SRE(dB) as a function of parameters λ and λ_{TV} for (Top row) DC1, (Middle row) DC2, and (Bottom row) DC3 with different SNRs.

data, we will be devoted to the self-adjustment of regularization coefficients as in [30], [31] in the future.

VI. REAL DATA EXPERIMENT

In this experiment, we verify the effectiveness of RSSUn-TV using a 350×350 -pixel subset of the well-known Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Cuprite data set, which is from the Cuprite region of Nevada, USA. The spectrum contains 188 bands with a wavelength range of $0.4 - 2.5 \mu\text{m}$ and a spectral resolution of 10 nm . Here, the spectral library used is $\mathbf{A} \in \mathbb{R}^{188 \times 240}$ from the USGS library with the low SNR and the absorption phase of the water vapor are removed. The Tetracorder 4.4 software² was used to inspect the different minerals in the mine to obtain the distribution of each mineral. Here, the maximum number of iterations is set to 300. Similarly, as in [9], we set the regularization parameters $\lambda = 0.001$ for SUNSAL, CLSUnSAL, and CLSUnL0 and set $\lambda = \lambda_{TV} = 0.001$ for SUNSAL-TV and RSSUn-TV.

Fig. 12 shows a qualitative comparison among the fractional abundance maps of the three considered minerals (i.e., *Alunite*, *Chalcedony*, and *Muscovite*) in the hyperspectral scene.

²<http://speclab.cr.usgs.gov/ARTICLE/tetracorder>

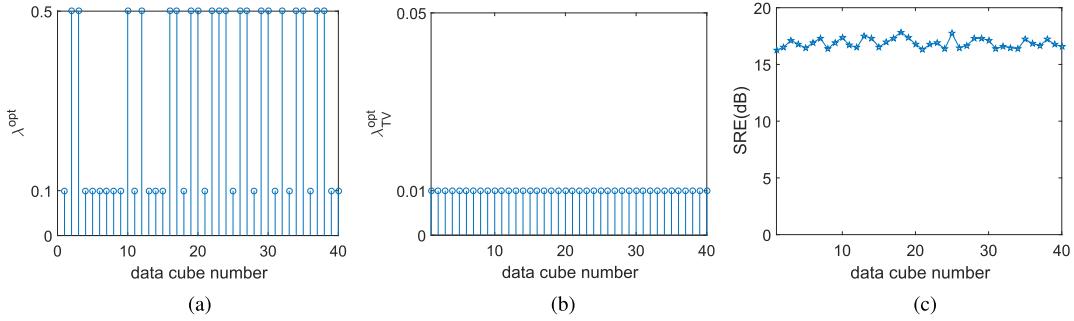


Fig. 11. Optimal parameters and SRE values of 40 different data cubes for DC1 with SNR=30 dB. (a) Value of optimal λ . (b) Value of optimal λ_{TV} . (c) Value of SRE.

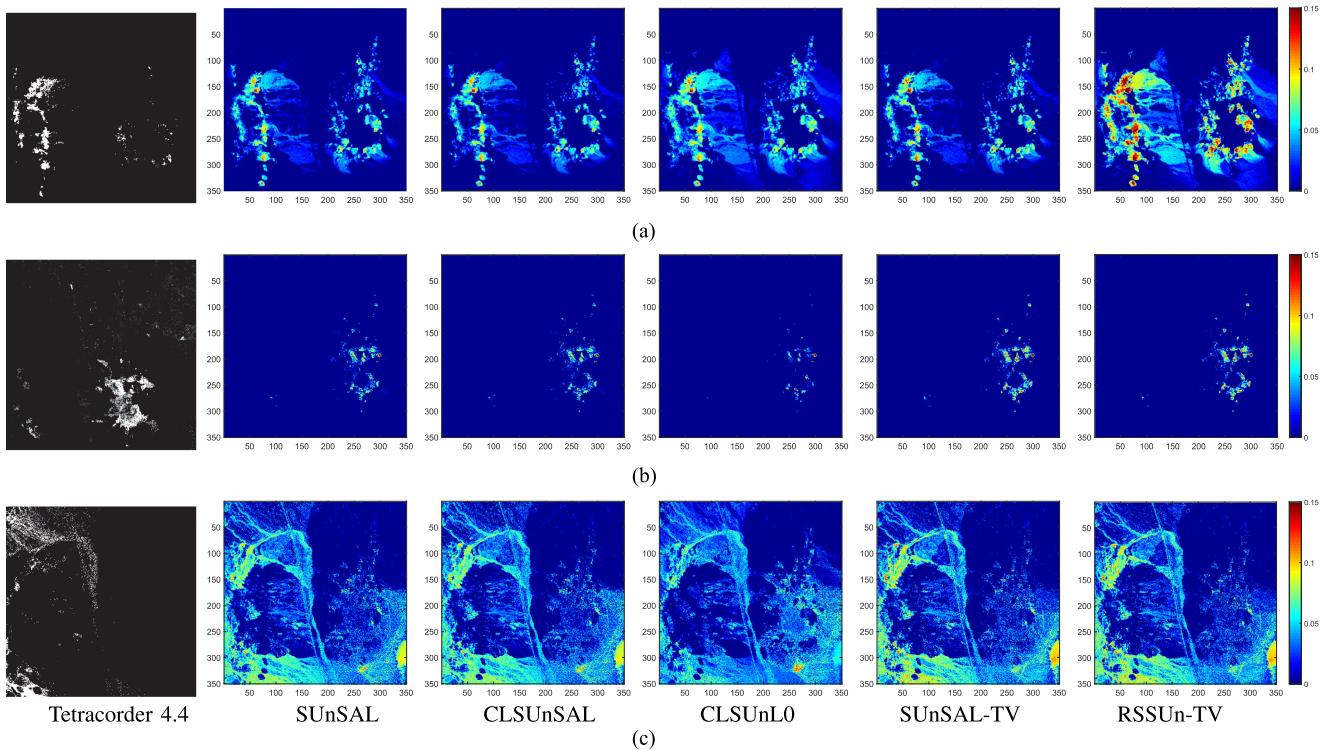


Fig. 12. Estimated abundance maps by Tetracorder 4.4 and different unmixing algorithms for the AVIRIS Cuprite scene for three minerals. (a) Alunite. (b) Chalcedony. (c) Muscovite.

Generally, the highest abundances estimated by the sparse unmixing algorithms correspond with those pixels belonging to the respective class of minerals. As evidenced from Fig. 12, five algorithms have achieved the acceptable unmixing results, similar to those of the Tetracorder software product. Another important observation from Fig. 12 is that the abundance maps estimated by the algorithms with the TV spatial regularizer present good spatial consistency of minerals and less low abundance values than others. Though the abundance maps estimated by SUNSAL-TV and RSSUn-TV appear very similar visually, the RSSUn-TV shows higher abundance estimate corresponding to Alunite. Therefore, we qualitatively conclude that RSSUn-TV offers promising results for the unmixing real hyperspectral data.

VII. CONCLUSION

In this article, we have proposed an RSSUn-TV. The RSSUn-TV algorithm couples the $\ell_{2,0}$ regularization with the TV regularization. Here, we use the $\ell_{2,0}$ regularization to promote the row sparsity in the abundance matrix. Considering that the $\ell_{2,0}$ regularization is sensitive to noise, the TV regularization in RSSUn-TV is used to suppress the noise and promote the piecewise smoothness of the abundance maps. We use the nonconvex ADMM to solve the proposed model and present some convergence analysis. Experimental results conducted on both simulated and real hyperspectral data sets indicate that the RSSUn-TV algorithm offers the potential to improve unmixing performance.

APPENDIX

A. Proof of Theorem 1

Proof: Here, we want to show the monotonicity of \mathcal{L}_μ in (10) by proving $\mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t+1)}, \mathbf{D}^{(t+1)}) \geq 0$. Specifically, we will first show that

$$\mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) \geq 0 \quad (13)$$

and show that

$$\mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t+1)}, \mathbf{D}^{(t+1)}) \geq 0. \quad (14)$$

To prove (13), we first show

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) \\ &= \frac{\mu}{2} \left\langle \mathbf{X}^{(t)} - \mathbf{X}^{(t+1)}, (\mathbf{A}^T \mathbf{A} + 2\mathbf{I})(\mathbf{X}^{(t)} + \mathbf{X}^{(t+1)}) \right. \\ &\quad \left. - 2(\mathbf{A}^T(\mathbf{V}_1^{(t)} + \mathbf{D}_1^{(t)})) - 2 \sum_{i=2,3} (\mathbf{V}_i^{(t)} + \mathbf{D}_i^{(t)}) \right\rangle \\ &= \frac{\mu}{2} \left\langle \mathbf{X}^{(t)} - \mathbf{X}^{(t+1)}, (\mathbf{A}^T \mathbf{A} + 2\mathbf{I})(\mathbf{X}^{(t)} - \mathbf{X}^{(t+1)}) \right\rangle \\ &\geq \frac{\mu}{2} \left\| \mathbf{X}^{(t)} - \mathbf{X}^{(t+1)} \right\|_F^2. \end{aligned}$$

To prove (14), we notice that $\mathbf{D}^{(t+1)} = \mathbf{D}^{(t)} - \mathbf{G}\mathbf{X}^{(t+1)} - \mathbf{B}\mathbf{V}^{(t+1)}$, i.e.,

$$\begin{cases} \mathbf{D}_1^{(t+1)} = \mathbf{D}_1^{(t)} - (\mathbf{A}\mathbf{X}^{(t+1)} - \mathbf{V}_1^{(t+1)}) \\ \mathbf{D}_2^{(t+1)} = \mathbf{D}_2^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_2^{(t+1)}) \\ \mathbf{D}_3^{(t+1)} = \mathbf{D}_3^{(t)} - (\mathbf{X}^{(t+1)} - \mathbf{V}_3^{(t+1)}) \\ \mathbf{D}_4^{(t+1)} = \mathbf{D}_4^{(t)} - (\mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t+1)}). \end{cases}$$

Then, we have

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t+1)}, \mathbf{D}^{(t+1)}) \\ &= g(\mathbf{V}^{(t)}) - g(\mathbf{V}^{(t+1)}) - \frac{\mu}{2} \sum_{i=1}^4 \left\| \mathbf{D}_i^{(t)} \right\|_F^2 \\ &\quad + \frac{\mu}{2} \sum_{i=1}^4 \left\| \mathbf{D}_i^{(t+1)} \right\|_F^2 \\ &\quad + \frac{\mu}{2} \left\| \mathbf{G}\mathbf{X}^{(t+1)} + \mathbf{B}\mathbf{V}^{(t)} - \mathbf{D}^{(t)} \right\|_F^2 \\ &\quad - \frac{\mu}{2} \left\| \mathbf{G}\mathbf{X}^{(t+1)} + \mathbf{B}\mathbf{V}^{(t+1)} - \mathbf{D}^{(t+1)} \right\|_F^2 \\ &= g(\mathbf{V}^{(t)}) - g(\mathbf{V}^{(t+1)}) - \frac{\mu}{2} \sum_{i=1}^4 \left\| \mathbf{D}_i^{(t)} \right\|_F^2 \\ &\quad + \frac{\mu}{2} \sum_{i=1}^4 \left\| \mathbf{D}_i^{(t+1)} \right\|_F^2 \\ &\quad + \frac{\mu}{2} \left\| \mathbf{B}(\mathbf{V}^{(t)} - \mathbf{V}^{(t+1)}) - \mathbf{D}^{(t+1)} \right\|_F^2 - \frac{\mu}{2} \left\| \mathbf{D}^{(t)} \right\|_F^2 \\ &\quad - 2 \left\| \mathbf{D}^{(t+1)} \right\|_F^2. \end{aligned}$$

In the following, we will discuss \mathcal{L}_μ with respect to \mathbf{V}_1 , \mathbf{V}_2 , \mathbf{V}_3 , and \mathbf{V}_4 separately. For simplicity, we let

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{V}_i^{(t)}, \mathbf{D}_i^{(t)}) &= \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}_{i-1}^{(t+1)}, \mathbf{V}_i^{(t)}, \mathbf{V}_{i+1}^{(t)}, \dots, \\ &\quad \mathbf{V}_4^{(t)}, \mathbf{D}_{i-1}^{(t)}, \mathbf{D}_i^{(t)}, \mathbf{D}_{i+1}^{(t)}, \dots, \mathbf{D}_5^{(t)}), \\ \mathcal{L}_\mu(\mathbf{V}_i^{(t+1)}, \mathbf{D}_i^{(t+1)}) &= \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}_{i-1}^{(t+1)}, \mathbf{V}_i^{(t+1)}, \mathbf{V}_{i+1}^{(t)}, \dots, \\ &\quad \mathbf{V}_4^{(t)}, \mathbf{D}_{i-1}^{(t)}, \mathbf{D}_i^{(t+1)}, \mathbf{D}_{i+1}^{(t)}, \dots, \mathbf{D}_5^{(t)}). \end{aligned}$$

Then, we want to prove the following inequality holds, i.e., $\mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}_i^{(t)}, \mathbf{D}_i^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}_i^{(t+1)}, \mathbf{D}_i^{(t+1)}) \geq 0$ for $i = 1, 2, 3, 4$. For \mathbf{V}_1 , we let $\mathcal{L}_\mu(\mathbf{V}_1) = \frac{1}{2} \|\mathbf{Y} - \mathbf{V}_1\|_F^2 + \frac{\mu}{2} \|\mathbf{AX} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2$. When updating \mathbf{V}_1 , the dual feasible condition about \mathbf{V}_1 is

$$\begin{aligned} \mathbf{0} &= \frac{\partial \mathcal{L}_\mu(\mathbf{V}_1)}{\partial \mathbf{V}_1} \\ &= \mathbf{V}_1^{(t+1)} - \mathbf{Y} + \mu \left(\mathbf{D}_1^{(t)} - (\mathbf{AX}^{(t+1)} - \mathbf{V}_1^{(t+1)}) \right) \\ &= \mathbf{V}_1^{(t+1)} - \mathbf{Y} + \mu \mathbf{D}_1^{(t+1)}. \end{aligned} \quad (15)$$

Then, by using the (16), we obtain the following inequality:

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{V}_1^{(t)}, \mathbf{D}_1^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_1^{(t+1)}, \mathbf{D}_1^{(t+1)}) \\ &= g_1(\mathbf{V}_1^{(t)}) - g_1(\mathbf{V}_1^{(t+1)}) + \frac{\mu}{2} \left\| \mathbf{D}_1^{(t+1)} \right\|_F^2 - \frac{\mu}{2} \left\| \mathbf{D}_1^{(t)} \right\|_F^2 \\ &\quad + \frac{\mu}{2} \left\| \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)} - \mathbf{D}_1^{(t+1)} \right\|_F^2 - \frac{\mu}{2} \left\| \mathbf{D}_1^{(t)} - 2\mathbf{D}_1^{(t+1)} \right\|_F^2 \\ &= \frac{1}{2} \left\langle \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)}, 2\mathbf{Y} - (\mathbf{V}_1^{(t+1)} + \mathbf{V}_1^{(t)}) \right\rangle \\ &\quad + \frac{\mu}{2} \left\| \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)} \right\|_F^2 - \frac{\mu}{2} \left\langle \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)}, 2\mathbf{D}_1^{(t+1)} \right\rangle \\ &\quad - \frac{\mu}{2} \left\| \mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)} \right\|_F^2 + \frac{\mu}{2} \left\langle \mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)}, 2\mathbf{D}_1^{(t+1)} \right\rangle \\ &\quad - \frac{\mu}{2} \left\| \mathbf{D}_1^{(t+1)} \right\|_F^2 + \frac{\mu}{2} \left\langle \mathbf{D}_1^{(t+1)} - \mathbf{D}_1^{(t)}, \mathbf{D}_1^{(t+1)} + \mathbf{D}_1^{(t)} \right\rangle \\ &= \frac{1}{2} \left\langle \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)}, \mu(\mathbf{D}_1^{(t+1)} + \mathbf{D}_1^{(t)}) \right\rangle \\ &\quad + \frac{\mu}{2} \left\| \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)} \right\|_F^2 - \mu \left\| \mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)} \right\|_F^2 \\ &\quad - \frac{\mu}{2} \left\langle \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)}, 2\mathbf{D}_1^{(t+1)} \right\rangle \\ &= \frac{1}{2} \left\langle \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)}, (\mu\mathbf{D}_1^{(t)} - \mathbf{Y}) - (\mu\mathbf{D}_1^{(t+1)} - \mathbf{Y}) \right\rangle \\ &\quad + \frac{\mu}{2} \left\| \mathbf{V}_1^{(t+1)} - \mathbf{V}_1^{(t)} \right\|_F^2 - \frac{1}{\mu} \left\| (\mu\mathbf{D}_1^{(t)} - \mathbf{Y}) - (\mu\mathbf{D}_1^{(t+1)} - \mathbf{Y}) \right\|_F^2 \\ &= \left(\frac{\mu+1}{2} - \frac{1}{\mu} \right) \left\| \mathbf{V}_1^{(t)} - \mathbf{V}_1^{(t+1)} \right\|_F^2. \end{aligned} \quad (16)$$

For \mathbf{V}_2 , we define $g_2(\mathbf{V}_2) = \lambda \|\mathbf{V}_2\|_{2,0}$. When updating \mathbf{V}_2 , we obtain

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{V}_2^{(t)}, \mathbf{D}_2^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_2^{(t+1)}, \mathbf{D}_2^{(t+1)}) \\ &= g_2(\mathbf{V}_2^{(t)}) - g_2(\mathbf{V}_2^{(t+1)}) - \frac{\mu}{2} \|\mathbf{D}_2^{(t)}\|_F^2 + \frac{\mu}{2} \|\mathbf{D}_2^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_2^{(t+1)} - \mathbf{V}_2^{(t)} - \mathbf{D}_2^{(t+1)}\|_F^2 \\ &\quad - \frac{\mu}{2} \|\mathbf{D}_2^{(t)} - \mathbf{D}_2^{(t+1)} - \mathbf{D}_2^{(t+1)}\|_F. \end{aligned} \quad (17)$$

We notice that the (17) with respect to \mathbf{V}_2 is same as the equality in [10]. Here, we cite the conclusion obtain by [10], so we have

$$\mathcal{L}_\mu(\mathbf{V}_2^{(t)}, \mathbf{D}_2^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_2^{(t+1)}, \mathbf{D}_2^{(t+1)}) \geq -8m\lambda \quad (18)$$

where m is the number of rows in \mathbf{V}_2 .

For \mathbf{V}_3 , we let

$$\mathcal{L}_\mu(\mathbf{V}_3) = \frac{\mu}{2} \|\mathbf{X} - \mathbf{V}_3 - \mathbf{D}_3\|_F^2 + \frac{\mu}{2} \|\mathbf{H}\mathbf{V}_3 - \mathbf{V}_4 - \mathbf{D}_4\|_F^2.$$

The dual feasible condition about \mathbf{V}_3 is

$$\begin{aligned} \mathbf{0} &= \frac{\partial \mathcal{L}_\mu}{\partial \mathbf{V}_3} \\ &= \mathbf{H}^T (\mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t)}) \\ &\quad - (\mathbf{X}^{(t+1)}\mathbf{V}_3^{(t+1)} - \mathbf{D}_3^{(t)}) \\ &= \mathbf{H}^T \mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{H}^T (\mathbf{V}_4^{(t)} + \mathbf{D}_4^{(t)}) + \mathbf{D}_3^{(t+1)}. \end{aligned} \quad (19)$$

Then, by using the (19), we obtain

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{V}_3^{(t)}, \mathbf{D}_3^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_3^{(t+1)}, \mathbf{D}_3^{(t+1)}) \\ &= \frac{\mu}{2} \|\mathbf{H}\mathbf{V}_3^{(t)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t)}\|_F^2 \\ &\quad - \frac{\mu}{2} \|\mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t)}\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_3^{(t+1)} - \mathbf{V}_3^{(t)} - \mathbf{D}_3^{(t+1)}\|_F^2 \\ &\quad - \frac{\mu}{2} \|\mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)} - \mathbf{D}_3^{(t+1)}\|_F^2 \\ &\quad - \frac{\mu}{2} \|\mathbf{D}_3^{(t)}\|_F^2 + \frac{\mu}{2} \|\mathbf{D}_3^{(t+1)}\|_F^2 \\ &= \frac{\mu}{2} \langle \mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}, \mathbf{H}^T \mathbf{H} (\mathbf{V}_3^{(t)} + \mathbf{V}_3^{(t+1)}) \rangle \\ &\quad - 2\mathbf{H}^T (\mathbf{V}_4^{(t)} + \mathbf{D}_4^{(t)}) \rangle + \frac{\mu}{2} \|\mathbf{V}_3^{(t+1)} - \mathbf{V}_3^{(t)}\|_F^2 \\ &\quad - \frac{\mu}{2} \langle \mathbf{V}_3^{(t+1)} - \mathbf{V}_3^{(t)}, 2\mathbf{D}_3^{(t+1)} \rangle - \frac{\mu}{2} \|\mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \langle \mathbf{D}_3^{(t+1)} - \mathbf{D}_3^{(t)}, \mathbf{D}_3^{(t+1)} + \mathbf{D}_3^{(t)} \rangle \\ &\quad + \frac{\mu}{2} \langle \mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)}, 2\mathbf{D}_3^{(t+1)} \rangle \\ &= \frac{\mu}{2} \langle \mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}, \mathbf{H}^T \mathbf{H} (\mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}) \rangle \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_3^{(t+1)} - \mathbf{V}_3^{(t)}\|_F^2 - \mu \|\mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)}\|_F^2. \end{aligned} \quad (20)$$

Recall that $\mathbf{D}_4^{(t+1)} = \mathbf{D}_4^{(t)} - (\mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t+1)})$, based on the dual feasible condition about \mathbf{V}_3 , we obtain

$$\begin{aligned} \mathbf{0} &= \frac{\partial \mathcal{L}_\mu}{\partial \mathbf{V}_3} \\ &= \mathbf{H}^T (\mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t)}) \\ &\quad - (\mathbf{X}^{(t+1)} - \mathbf{V}_3^{(t+1)} - \mathbf{D}_3^{(t)}) \\ &= \mathbf{H}^T (\mathbf{V}_4^{(t+1)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t+1)}) + \mathbf{D}_3^{(t+1)}. \end{aligned}$$

Thus, $\mathbf{D}_3^{(t+1)} = \mathbf{H}^T (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) + \mathbf{H}^T \mathbf{D}_4^{(t+1)}$. By using the triangle inequality, we obtain

$$\begin{aligned} & \|\mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)}\|_F^2 \\ &\leq 2 \left(\langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, \mathbf{H}^T \mathbf{H} (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle \right. \\ &\quad \left. + \|\mathbf{H}\|_F^2 \|\mathbf{D}_4^{(t+1)}\|_F^2 \right). \end{aligned} \quad (21)$$

For \mathbf{V}_4 , let $\mathcal{L}_\mu(\mathbf{V}_4) = \lambda_{TV} \|\mathbf{V}_4\|_{1,1} + \frac{\mu}{2} \|\mathbf{H}\mathbf{V}_3 - \mathbf{V}_4 - \mathbf{D}_4\|_F^2$ and define $g_4(\mathbf{V}_4) = \lambda_{TV} \|\mathbf{V}_4\|_{1,1}$. When updating \mathbf{V}_4 , we write the dual feasible condition about \mathbf{V}_4 as

$$\begin{aligned} \mathbf{0} &\in \frac{\partial \mathcal{L}_\mu}{\partial \mathbf{V}_4} \\ &= \nabla g_4(\mathbf{V}_4^{(t+1)}) - \mu (-\mathbf{D}_4^{(t)} + \mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{V}_4^{(t+1)}) \\ &= \nabla g_4(\mathbf{V}_4^{(t+1)}) + \mu \mathbf{D}_4^{(t+1)}. \end{aligned} \quad (22)$$

Then, by using the (22), we have

$$\begin{aligned} & \mathcal{L}_\mu(\mathbf{V}_4^{(t)}, \mathbf{D}_4^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_4^{(t+1)}, \mathbf{D}_4^{(t+1)}) \\ &= g_4(\mathbf{V}_4^{(t)}) - g_4(\mathbf{V}_4^{(t+1)}) - \frac{\mu}{2} \|\mathbf{D}_4^{(t)}\|_F^2 + \frac{\mu}{2} \|\mathbf{D}_4^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_4^{(t+1)} - \mathbf{V}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2 \\ &\quad - \frac{\mu}{2} \|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)} - \mathbf{D}_4^{(t+1)}\|_F^2 \\ &= g_4(\mathbf{V}_4^{(t)}) - g_4(\mathbf{V}_4^{(t+1)}) + \frac{\mu}{2} \|\mathbf{V}_4^{(t+1)} - \mathbf{V}_4^{(t)}\|_F^2 \\ &\quad - \frac{\mu}{2} \langle \mathbf{V}_4^{(t+1)} - \mathbf{V}_4^{(t)}, 2\mathbf{D}_4^{(t+1)} \rangle - \frac{\mu}{2} \|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \langle \mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}, 2\mathbf{D}_4^{(t+1)} \rangle \\ &\quad + \frac{\mu}{2} \langle \mathbf{D}_4^{(t+1)} - \mathbf{D}_4^{(t)}, \mathbf{D}_4^{(t+1)} + \mathbf{D}_4^{(t)} \rangle \\ &= g_4(\mathbf{V}_4^{(t)}) - g_4(\mathbf{V}_4^{(t+1)}) + \frac{\mu}{2} \|\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}\|_F^2 \\ &\quad - \frac{1}{2} \langle \mathbf{V}_4^{(t+1)} - \mathbf{V}_4^{(t)}, \nabla g_4(\mathbf{V}_4^{(t+1)}) \rangle \\ &\quad - \mu \|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2 \\ &\geq \frac{\mu}{2} \|\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}\|_F^2 - \mu \|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2. \end{aligned} \quad (23)$$

The last inequality holds due to the convexity of $g_4(\mathbf{V}_4) = \lambda_{TV} \|\mathbf{V}_4\|_{1,1}$. Besides, based on the *soft-threshold* function in [24], [25], we let $\varpi = \mathbf{H}\mathbf{V}_3^{(t+1)} - \mathbf{D}_4^{(t)}$, and elicit the following equations:

$$\begin{aligned}\mathbf{V}_4^{(t+1)} &= \text{soft} \left(\mathbf{D}_4^{(t)} - \mathbf{H}\mathbf{V}_3^{(t+1)}, \frac{\lambda_{TV}}{\mu} \right) \\ &= \begin{cases} \varpi + \frac{\lambda_{TV}}{\mu}, & \varpi < -\frac{\lambda_{TV}}{\mu} \\ \mathbf{0}, & |\varpi| < \frac{\lambda_{TV}}{\mu} \\ \varpi - \frac{\lambda_{TV}}{\mu}, & \varpi > \frac{\lambda_{TV}}{\mu}. \end{cases}\end{aligned}$$

It follows that

$$\begin{aligned}\mathbf{D}_4^{(t+1)} &= \mathbf{D}_4^{(t)} - \mathbf{H}\mathbf{V}_3^{(t+1)} + \mathbf{V}_4^{(t+1)} \\ &= \begin{cases} \frac{\lambda_{TV}}{\mu}, & \varpi < -\frac{\lambda_{TV}}{\mu} \\ \mathbf{D}_4^{(t)} - \mathbf{H}\mathbf{V}_3^{(t+1)}, & |\varpi| < \frac{\lambda_{TV}}{\mu} \\ -\frac{\lambda_{TV}}{\mu}, & \varpi > \frac{\lambda_{TV}}{\mu}. \end{cases}\end{aligned}$$

That is to say, $\|\mathbf{D}_4^{(t+1)}\|_F^2 \leq (\frac{\lambda_{TV}}{\mu})^2$. Therefore, combined with (23), we have

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{V}_4^{(t)}, \mathbf{D}_4^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_4^{(t+1)}, \mathbf{D}_4^{(t+1)}) &\geq \frac{\mu}{2} \|\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}\|_F^2 - 2 \frac{\lambda_{TV}^2}{\mu}.\end{aligned}$$

Moreover, for (21), we let $\hbar = \|\mathbf{H}\|_F^2$ and obtain

$$\begin{aligned}\|\mathbf{D}_3^{(t)} - \mathbf{D}_3^{(t+1)}\|_F^2 &\leq 2 \left(\langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, \mathbf{H}^T \mathbf{H} \times (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle + \hbar \|\mathbf{D}_4^{(t+1)}\|_F^2 \right) \\ &\leq 2 \left(\langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, \mathbf{H}^T \mathbf{H} \times (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle + \left(\hbar \frac{\lambda_{TV}}{\mu} \right)^2 \right).\end{aligned}$$

Thus, for (20), we have

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{V}_3^{(t)}, \mathbf{D}_3^{(t)}) - \mathcal{L}_\mu(\mathbf{V}_3^{(t+1)}, \mathbf{D}_3^{(t+1)}) &\geq \frac{\mu}{2} \langle \mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}, \mathbf{H}^T \mathbf{H} (\mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}) \rangle \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_3^{(t+1)} - \mathbf{V}_3^{(t)}\|_F^2 \\ &\quad - 2\mu \left(\langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, \mathbf{H}^T \mathbf{H} (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle \right. \\ &\quad \left. + \hbar \left(\frac{\lambda_{TV}}{\mu} \right)^2 \right).\end{aligned}\tag{24}$$

As a conclusion, from (13), (16), (18), (24), and (23), the change of the original Lagrange function during an iteration

cycle can be expressed as follows:

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) - \mathcal{L}_\mu(\mathbf{X}^{(t+1)}, \mathbf{V}^{(t+1)}, \mathbf{D}^{(t+1)}) &\geq \frac{\mu}{2} \|\mathbf{X}^{(t)} - \mathbf{X}^{(t+1)}\|_F^2 + \frac{\mu}{2} \|\mathbf{V}_1^{(t)} - \mathbf{V}_1^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \langle \mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}, (\mathbf{I} + \mathbf{H}^T \mathbf{H}) (\mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}) \rangle \\ &\quad + \frac{\mu}{2} \langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, (\mathbf{I} - 4\mathbf{H}^T \mathbf{H}) (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle \\ &\quad - 8m\lambda - 2(1 + \hbar) \frac{\lambda_{TV}^2}{\mu} \\ &= h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) + \mathbf{C}\end{aligned}\tag{25}$$

where

$$\begin{aligned}h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) &= \frac{\mu}{2} \|\mathbf{X}^{(t)} - \mathbf{X}^{(t+1)}\|_F^2 + \frac{\mu}{2} \|\mathbf{V}_1^{(t)} - \mathbf{V}_1^{(t+1)}\|_F^2 \\ &\quad + \frac{\mu}{2} \langle \mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}, (\mathbf{I} + \mathbf{H}^T \mathbf{H}) (\mathbf{V}_3^{(t)} - \mathbf{V}_3^{(t+1)}) \rangle \\ &\quad + \frac{\mu}{2} \langle \mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}, (\mathbf{I} - 4\mathbf{H}^T \mathbf{H}) (\mathbf{V}_4^{(t)} - \mathbf{V}_4^{(t+1)}) \rangle\end{aligned}$$

and

$$\mathbf{C} = 8m\lambda + 2(1 + \hbar) \frac{\lambda_{TV}^2}{\mu}.\tag{26}$$

That is to say, when $\mathbf{C} \leq h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)})$ holds, the \mathcal{L}_μ in (10) is decreasing.

In the (26), we only give a dynamic range for $\lambda, \lambda_{TV}, \mu$, and then we try to find a fixed range for them. Next, we will find the lower bound of $h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)})$, which is the fixed range. First, we relax $h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)})$ as follows:

$$\begin{aligned}h(\mathbf{X}^{(t)}, \mathbf{V}^{(t)}, \mathbf{D}^{(t)}) &\geq \frac{\mu}{2} \|\mathbf{V}_1^{(t)} - \mathbf{V}_1^{(t+1)}\|_F^2 \\ &= \frac{\mu^3}{2} \|\mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)}\|_F^2.\end{aligned}$$

So, the following inequality promise that (25) is always negative:

$$\mathbf{C} \leq \frac{\mu^3}{2} \|\mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)}\|_F^2.$$

Based on the row-hard-threshold function, we assume that the row vectors $(\mathbf{v}_2)_i, (\mathbf{d}_2)_i$, and $(\mathbf{x})_i$ are the i th row in $\mathbf{V}_2, \mathbf{D}_2$, and \mathbf{X} and magnify the value of $\|\mathbf{D}_2^{(t)} - \mathbf{D}_2^{(t+1)}\|_F^2$. Recall that

$$\mathbf{V}_2^{(t+1)} = \mathbf{T}_{row}^{VHT} \left(\mathbf{X}^{(t+1)} - \mathbf{D}_2^{(t)}, \sqrt{\frac{2\lambda}{\mu}} \right)$$

It leads to

$$\begin{aligned} \left(\mathbf{d}_2^{(t+1)}\right)_i &= \left(\mathbf{d}_2^{(t)}\right)_i - (\mathbf{x}^{(t+1)})_i + \left(\mathbf{v}_2^{(t+1)}\right)_i \\ &= \begin{cases} \mathbf{0}, & \|(\mathbf{x}^{(t+1)})_i - (\mathbf{d}_2^{(t)})_i\|_2 > \sqrt{\frac{2\lambda}{\mu}} \\ (\mathbf{d}_2^{(t)})_i - (\mathbf{x}^{(t+1)})_i, & \|(\mathbf{x}^{(t+1)})_i - (\mathbf{d}_2^{(t)})_i\|_2 \leq \sqrt{\frac{2\lambda}{\mu}} \end{cases}. \end{aligned}$$

That is to say, $\|\mathbf{D}_2^{(t+1)}\|_F^2 \leq \frac{2\lambda}{\mu}$. Therefore, the following inequality can be obtained by using the triangle inequality:

$$\|\mathbf{D}_2^{(t)} - \mathbf{D}_2^{(t+1)}\|_F^2 \leq \frac{4\lambda}{\mu}.$$

Similarly, we magnify the value of $\|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2$:

$$\|\mathbf{D}_4^{(t)} - \mathbf{D}_4^{(t+1)}\|_F^2 \leq 2 \left(\frac{\lambda_{TV}}{\mu} \right)^2.$$

When the algorithm does not reach the stopping criterion, which means $\|\mathbf{GX}^{(t)} + \mathbf{BV}^{(t)}\|_F \geq \varepsilon$. Then, we have

$$\|\mathbf{D}_1^{(t)} - \mathbf{D}_1^{(t+1)}\|_F^2 \leq \left(\varepsilon - 2 \left(\frac{\lambda_{TV}}{\mu} \right)^2 - \frac{4\lambda}{\mu} \right)^2.$$

Finally, combining the above formulas, we obtain

$$\mathbf{C} \leq \frac{\mu^3}{2} \left(\varepsilon - 2 \left(\frac{\lambda_{TV}}{\mu} \right)^2 - \frac{4\lambda}{\mu} \right)^2. \quad (27)$$

■

REFERENCES

- [1] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [2] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
- [3] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [4] J. Huang, T. Z. Huang, L. J. Deng, and X. L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2438, Apr. 2019.
- [5] J. Huang, T. Z. Huang, X. L. Zhao, and L. J. Deng, "Joint-sparse-blocks regression for total variation regularized hyperspectral unmixing," *IEEE Access*, vol. 7, pp. 138779–138791, 2019.
- [6] X. Li, J. Huang, L. J. Deng, and T. Z. Huang, "Bilateral filter based total variation regularization for sparse hyperspectral image unmixing," *Inf. Sci.*, vol. 504, pp. 334–353, 2019.
- [7] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [8] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
- [9] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [10] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using ℓ_0 norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5495–5508, Sep. 2018.
- [11] Z. Cui, H. Zhang, and W. Lu, "An improved smoothed ℓ_0 -norm algorithm based on multiparameter approximation function," *Int. Conf. Commun. Technol.*, pp. 942–945, 2010.
- [12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenom.*, vol. 60, pp. 259–268, 1992.
- [13] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.
- [14] H. Zhang, "Hyperspectral image denoising with cubic total variation model," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. I-7, pp. 95–98, 2012.
- [15] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4257–4266, Sep. 2016.
- [16] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.
- [17] X. L. Zhao, F. Wang, T. Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.
- [18] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 529–545, Mar. 2001.
- [19] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
- [20] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [21] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [22] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.
- [23] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in non-convex nonsmooth optimization," *J. Sci. Comput.*, pp. 29–63, 2018.
- [24] J. F. Cai, S. Osher, and Z. Shen, "Convergence of the linearized Bregman iteration for ℓ_p -norm minimization," *Math. Comput.*, vol. 78, no. 268, pp. 2127–2136, 2009.
- [25] W. Yin, D. G. S. Osher, and J. Darbon, "Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 143–168, 2008.
- [26] K. Abercromby, "Communication of the NASA JSC spacecraft materials spectral database," 2006.
- [27] L. F., M. Ng, and R. Plemmons, "Coupled segmentation and denoising/deblurring models for hyperspectral material identification," *Numer. Linear Algebra Appl.*, vol. 19, no. 1, pp. 153–173, 2012.
- [28] R. P. Q. Zhang, H. Wang, and V. Pauca, "Tensor methods for hyperspectral data analysis: A space object material identification study," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 25, no. 12, pp. 3001–3012, 2008.
- [29] M. G. T. Blake, S. Cain, and K. Jerkatis, "Reconstruction of spectral images from the AEOS spectral imaging sensor," in *Proc. Adv. Maui Opt. Space Surveillance Tech. Conf.*, 2006, pp. E1–E22.
- [30] X. Xia and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 54–69, 2017.
- [31] X. Xia, Z. Shi, and B. Pan, " ℓ_0 -based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 46–58, 2018.