Spectral-Spatial Constrained Nonnegative Matrix Factorization for Spectral Mixture Analysis of Hyperspectral Images

Ge Zhang, *Student Member, IEEE*, Shaohui Mei, *Senior Member, IEEE*, Yan Feng, and Qian Du, *Fellow, IEEE*

***Abstract*—Hyperspectral spectral mixture analysis (SMA), which intends to decompose mixed pixels into a collection of endmembers weighted by their corresponding fraction abundances, has been successfully used to tackle mixed-pixel problem in hyperspectral remote sensing applications. As an approach of decomposing a high-dimensional data matrix into the multiplication of two nonnegative matrices, nonnegative matrix factorization (NMF) has shown its advantages and been widely applied to SMA. Unfortunately, most of the NMF-based unmixing methods can easily lead to an unsuitable solution, due to inadequate mining of spatial and spectral information and the influence of outliers and noise. To overcome such limitations, a spatial constraint over abundance and a spectral constraint over endmembers are imposed over NMF-based unmixing model for spectral-spatial constrained unmixing. Specifically, a spatial neighborhood preserving constraint is proposed to preserve the local geometric structure of the hyperspectral data by assuming that pixels in a spatial neighborhood generally fall into a low-dimensional manifold, while a minimum spectral distance constraint is formulated to optimize endmember spectra as compact as possible. Furthermore, to handle non-Gaussian noises or outliers, an $L_{2,1}$-norm based loss function is ultimately adopted for the proposed spectral-spatial constrained nonnegative matrix factorization model and a projected gradient based optimization algorithm is designed for optimization. Experimental results over both synthetic and real-world datasets demonstrate that the proposed spatial and spectral constraints can certainly improve the performance of NMF-based unmixing and outperform state-of-the-art NMF-based unmixing algorithms.**

***Index Terms*—Hyperspectral remote sensing, linear mixture model (LMM), nonnegative matrix factorization (NMF), spectral mixture analysis (SMA).**

I. Introduction

HYPERSPECTRAL remote sensing image, which involves collecting abundant spectral information over hundreds of contiguous bands, has been widely applied to many civil and military fields [1]–[3]. However, in view of the low spatial resolution of imaging system, mixed pixels are commonly found in hyperspectral images [4]–[6]. The presence of mixed pixels seriously affects the performance of many image processing tasks, such as classification and target detection, limiting the quantitative development of hyperspectral remote sensing. Therefore, spectral mixture analysis (SMA), which extracts fractional abundance for pure ground objects (known as endmembers) within a pixel, has been developed to solve the mixed-pixel problems for quantitative analysis of hyperspectral remote sensing images. In recent years, the development of deep learning technology has promoted the development of hyperspectral unmixing [7]–[9], but it is still limited since large amount of training samples are required.

In the past decades, many unsupervised SMA algorithms have been proposed by jointly performing endmember extraction and abundance estimation, such as unconstrained fully constrained least square unmixing algorithm [10], gradient descent maximum entropy algorithm [11], to name a few. In these algorithms, pure pixels for each endmember are assumed to present in the data. However, such a strong requirement is not met when pixels are in a high mixing level. Therefore, a group of algorithms have been developed for highly mixed situation by considering the unsupervised SMA as a blind source separation problem [12].

Independent component analysis (ICA), which extracts independent sources from observed data, has been applied to SMA by treating it as a blind source separation problem [13]. However, it has been demonstrated that ICA is not a suitable approach since both endmembers and abundances are not statistically independent, which violates the assumption of ICA. Meanwhile, nonnegativity of endmembers and abundances cannot be guaranteed in ICA. To overcome these limitations of ICA, Nascimento *et al.* [14] proposed the dependent component analysis method based on the assumption that the columns of abundance matrix are random vectors following a mixture of Dirichlet probability density functions. However, nonnegativity of endmember spectra is still not guaranteed and may even result in "negative estimated reflectance."

Nonnegative matrix factorization (NMF), which was proposed to find a set of nonnegative basis functions for representing nonnegative data [15], has been widely utilized for SMA of highly mixed data since the nonnegativity and sum-to-one

Manuscript received February 8, 2021; revised March 21, 2021; accepted June 17, 2021. Date of publication June 25, 2021; date of current version July 14, 2021. This work was supported in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX2021080, in part by the Fundamental Research Funds for the Central Universities, and in part by the National Natural Science Foundation of China under Grant 61671383. *(Corresponding author: Shaohui Mei.)*

Ge Zhang, Shaohui Mei, and Yan Feng are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: zhangg@mail.nwpu.edu.cn; meish@nwpu.edu.cn; sycfy@nwpu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/JSTARS.2021.3092566

constraints can be easily relaxed [16], [17]. However, minimizing the representation error in the linear mixture model (LMM) by NMF is not sufficient for SMA since the unmixing results of NMF are not unique [18]. In order to make this problem well-posed, auxiliary constraints must be imposed to NMF for physically meaningful results. Many constrained NMF algorithms have been proposed for this purpose, which can be classified into three categories, i.e., endmember constrained NMF, abundance constrained NMF, and double constrained NMF.

One of the widely used endmenber constraints is variational minimum volume. Given that image pixels are generally assumed to lie in a simplex whose vertices are endmembers [19], the minimum volume constrained NMF (MVC-NMF) algorithm was proposed to restrict the volume of the simplex when NMF is employed for SMA [20]. The volume-regularized NMF algorithm further promoted small volume of the convex hull spanned by the basis matrix [21]. Zhuang *et al.* [22] discussed the use of several minimum volume regularizations and introduced a strategy named NMF-QMV for automatically selecting regularization parameter, which included by different regularization modules based on boundary, center, and total variation (TV). However, matrix determinant and matrix inversion computation may cause numerical instability problems when the simplex volume becomes small. The minimum dispersion constrained NMF (MiniDisCo-NMF) was also proposed for SMA such that the endmember spectra have minimum variances, while it may oversmooth the endmember spectra [23]. A similar approach was also proposed to constrain endmember dissimilarity, which can not only measure the difference between the endmember signatures but also constrain the obtained signatures to be smooth [24]. In addition, prior knowledge over endmembers, such as spectral signature of known endmembers, can be propagated during optimization process by minimizing the difference between the image data and the prior knowledge [25].

Recently, abundance constrained NMF algorithms have gained great attention in SMA of hyperspectral images. The smoothness constraint was imposed over abundance, which can be realized by the TV regularizer, following the assumption that ground objects usually vary slowly and abrupt changes rarely appear in an image [26], [27]. Moreover, since the number of endmembers in an entire image is usually much larger than the number of endmembers present in a specific pixel, especially in the spectral library based unmixing, it often leads to a sparse solution and thus it is intuitive and meaningful to constrain NMF with sparseness [28]. Qian *et al.* utilized an $L_{1/2}$ norm to ensure the sparsity of abundance [29]. Feng *et al.* proposed a sparsity-constrained deep NMF with the total variation (SDNMF-TV) technique for hyperspectral unmixing, which introduced an $L_{1/2}$ constraint for the sparse distribution of each endmember in the 2-D space and the TV regularizer for promoting piecewise smoothness in abundance maps [30]. However, the sparsity constraint is unstable and prone to noise. Spatial context, which affects the abundance relationship between pixels and their neighbors, has also attracted much attention in SMA applications [31]. Specifically, the abundance structural constraint was adopted since spectra in the homogeneous region are considered to express similar substance constructions and the same sparsity [32], [33]. The region-based structure preserving NMF explored consistent data distribution in the same region while discriminating different data structures across regions in the unmixed data [34]. Wang *et al.* proposed the spatial group sparsity-regularized NMF (SGSNMF) [35], but it only accounted for spatial homogeneity of the first-order pixel neighborhood system of the abundance map, the spatial structure information of the image remained incomplete. Yang *et al.* combined nonlocal spatial information with spatial group sparsity together into the NMF method [36].

Taking either endmember or abundance constraint into account in NMF based unmixing has a restricted effect to obtain a satisfying solution. Therefore, double-constrained NMF algorithms are expected to impose constrains on both endmembers and abundance simultaneously. Jia *et al.* proposed the piecewise smooth NMF with sparseness constraint algorithm by incorporating a united constraints of the piecewise smoothness and the sparseness [37]. However, it often needs prior knowledge about the exact sparseness degree of abundances. Salehani *et al.* explored the use of "arctan" function as a sparse regularizer term, and enforced a roughness penalty to make spectral signatures smooth as well as to restrict the related estimation error [38]. The double-constrained multilayer NMF method [39] adopted the sparsity of the abundance matrix and a corresponding graph regularization term based on the *k*-nearest neighbor graph to preserve the geometric structure of the data. The robust collaborative nonnegative matrix factorization (R-CoNMF) [40] addressed sparsity among the lines of the abundance matrix and promoted minimum volume by pushing the endmembers toward the mean value of the data set or bring the endmembers to the real solution quadratically regularized by a given simplex. Yuan *et al.* further introduced TV regularization into R-CoNMF to better handle the huge solution space problem and explore the adjacent relationship [41]. Specially, simplex minimum volume, abundance sparsity, and abundance smoothness have been simultaneously studied in the multiple-priors ensemble constrained NMF model [42]. Many other related methods, such as graph regularized based [43], self-paced based NMF algorithms [44], tensor based [45], [46], and deep learning based unmixing algorithms [47]–[49], have also been developed. However, these algorithms do not deal well with outliers or noise and the effect on spatial structure of abundance distribution still needs to be further verified.

To alleviate the above-mentioned problems and take full advantage of spatial and spectral information, in this article, a spectral-spatial constrained NMF algorithm is proposed by imposing constraints on both endmembers and abundance simultaneously. Specifically, a spatial neighborhood preserving constraint is imposed on abundances to preserve local geometric structure of the data in a low-dimensional manifold. Meanwhile, a minimum spectral distance constraint is also introduced to optimize endmember spectra as compact as possible. To further handle non-Gaussian noise or outliers, an $L_{2,1}$-norm based loss function is adopted for the proposed spectral-spatial constrained nonnegative matrix factorization (SS-NMF) algorithm. Finally, extensive experiments over both synthetic and real-world datasets are conducted to validate the effectiveness and

robustness of the proposed SS-NMF methods for hyperspectral unmixing.

The rest of this article is organized as follows. Section II introduces the basic idea of NMF algorithms. Section III presents the proposed SS-NMF model for SMA of hyperspectral images. Section IV reports and discusses experimental results on both synthetic and real datasets. Finally, Section V concludes this article.

II. NMF Algorithms for SMA

A. NMF Model for SMA

The LMM, which is well-known for its simplicity and explicit physical meaning, has been widely utilized to model the relationship between mixtures and endmembers. In the LMM, the photons reflected from different ground objects within one pixel are assumed not to interfere with each other. As a result, the observed spectral matrix $\mathbf{R} \in \mathcal{R}^{b\times o}$ is assumed to be the product of endmember matrix $\mathbf{M} \in \mathcal{R}^{b\times c}$ and its corresponding abundance matrix $\mathbf{A} \in \mathcal{R}^{c\times o}$ plus a noise matrix $\mathbf{N} \in \mathcal{R}^{b\times o}$. Their relationship can be formulated as

$$\mathbf{R} = \mathbf{MA} + \mathbf{N} \tag{1}$$

in which b is the number of bands, c is the number of endmembers, and o is the number of pixels in the image. Here, the ith column vectors of $\mathbf{M}$ and $\mathbf{A}$, denoted by $\mathbf{m}_i$ and $\mathbf{a}_i$, correspond to the ith endmember and the abundance of the ith spectral pixel $\mathbf{r}_i$, respectively.

In order to make the LMM physically meaningful, the abundance sum-to-one constraint (ASC) and the abundance nonnegative constraint (ANC) must be imposed, which can be expressed as

$$\sum_{j=1}^{c} A_{jk} = 1 \tag{2}$$

$$A_{jk} \geq 0 \qquad j = 1, 2, \ldots, c, k = 1, 2, \ldots, o. \tag{3}$$

Generally, the ASC can be embedded into this model by adding an additional pseudoband to the data matrix and the endmember matrix [50]. Therefore, in the following analysis, only the ANC is explicitly considered.

Given the target matrix $\mathbf{R}$, the task of NMF is to decompose it into two nonnegative matrices $\mathbf{M}$ and $\mathbf{A}$. Hence, the following model can be utilized to tackle the mixed-pixel problem:

$$\mathbf{R} \approx \mathbf{MA}$$
$$\text{s.t.} \mathbf{M} \succeq 0, \mathbf{A} \succeq 0 \tag{4}$$

where "$\succeq 0$" represents elementwise nonnegativity. Therefore, the generally used Frobenius-norm based NMF model for SMA can be expressed by the following optimization problem:

$$\min \frac{1}{2}\|\mathbf{R} - \mathbf{MA}\|_F$$
$$\text{s.t.} \mathbf{M} \succeq 0, \mathbf{A} \succeq 0. \tag{5}$$

Since a real world hyperspectral data contain outliers and noise, especially non-Gaussian noise may exist, the errors for both pixels and samples are inevitably squared, which means a few pixels that contain noise or outliers with large error values will play a major role in the objective function [51]. To overcome the above-mentioned problems, we employ joint an $L_{2,1}$-norm loss function to resist the sensitivity to noises and outliers, which has been shown to be very effective in [52]. Therefore, the NMF optimization problem based on the $L_{2,1}$-norm loss function can be formulated as

$$\min f(\mathbf{M}, \mathbf{A}) = \frac{1}{2}\|\mathbf{R} - \mathbf{MA}\|_{2,1} = \frac{1}{2}\sum_{i=1}^{b} \|\mathbf{R}_i - (\mathbf{MA})_i\|_2$$
$$\text{s.t.} \mathbf{M} \succeq 0, \mathbf{A} \succeq 0 \tag{6}$$

where $\mathbf{R}_i$ and $(\mathbf{MA})_i$ are the ith row of $\mathbf{R}$ and $\mathbf{MA}$, respectively. In addition, according to [53], the gradient derivations of each variable can be determined as follows:

$$\nabla_{\mathbf{M}} f(\mathbf{M}, \mathbf{A}) = \mathbf{Q}(\mathbf{MA} - \mathbf{R})\mathbf{A}^T \tag{7}$$

$$\nabla_{\mathbf{A}} f(\mathbf{M}, \mathbf{A}) = \mathbf{M}^T\mathbf{Q}(\mathbf{MA} - \mathbf{R}) \tag{8}$$

where $\mathbf{Q} \in \mathcal{R}^{b\times b}$ is a diagonal matrix with the ith diagonal element as

$$q_{ii} = \frac{1}{2\|\mathbf{R}_i - (\mathbf{MA})_i\|_2}. \tag{9}$$

B. PG Based NMF Algorithm for SMA

Many algorithms have been proposed to solve the objective function of NMF problem, such as multiplicative updating method [54], projected gradient (PG) method [55], active set method [56], and block principal pivoting method [57], in which PG based algorithm has been demonstrated to be very effective to solve the constrained NMF problem for SMA. In order to minimize the cost function of NMF [defined by (6)), the PG method is utilized to solve two subproblems of NMF alternatively, which is summarized in Algorithm 1.

Algorithm 1: The alternating PG algorithm for NMF.

1. Initializing $\mathbf{M}^1 \succeq 0, \mathbf{A}^1 \succeq 0$.
2. For $k = 1, 2, \ldots$

$$\mathbf{M}^{k+1} = P\left[\mathbf{M}^k - \alpha_k \nabla_{\mathbf{M}} f(\mathbf{M}^k, \mathbf{A}^k)\right] \tag{10}$$

$$\mathbf{A}^{k+1} = P\left[\mathbf{A}^k - \alpha_k \nabla_{\mathbf{A}} f(\mathbf{M}^{k+1}, \mathbf{A}^k)\right] \tag{11}$$

where $P[x]$ is a projected function, the value is chosen as follows:

$$P[x] = \begin{cases} 0 \text{ if } x \leqslant 0 \\ x \text{ if } 0 < x < u, \\ u \text{ if } x \geqslant u \end{cases} \tag{12}$$

in which u is the upper bound for $\mathbf{M}$ or $\mathbf{A}$.

One of the key problems in the PG method is selecting the step size α_k, which is also the most time consuming for this algorithm. An simple and effective solution to select α_k is "Armijo rule along the projection arc" [58].

III. Our Proposed Spectral and Spatial Constrained NMF Algorithms for SMA

It has been demonstrated that minimizing the representation error in LMM by NMF is not sufficient for SMA since the unmixing results of NMF are not unique. In order to make this problem well-posed, extra constraints other than the ANC must be imposed to the objective function in (6)

$$\mathbf{F}(\mathbf{M},\mathbf{A}) = \frac{1}{2}\|\mathbf{R}-\mathbf{MA}\|_{2,1} + \frac{1}{2}\lambda_1 \cdot J_1(\mathbf{M}) + \frac{1}{2}\lambda_2 \cdot J_2(\mathbf{A}), \tag{13}$$

in which $J_1(\mathbf{M})$ and $J_2(\mathbf{A})$ are additional constraints imposed on endmember matrix $\mathbf{M}$ and abundance matrix $\mathbf{A}$, respectively, and λ_1 and λ_2 are their tradeoffs among different objective functions.

Observed from the LMM defined by (1), spectral information of an image can be represented by its endmembers since all the pixels can be represented by these endmembers, while spatial information, which reflects the structural relationship of pixels and their neighbors, can be expressed in the similarity of their corresponding abundance. Therefore, in this article, the spectral constraint is considered for $J_1(\mathbf{M})$ and the spatial constraint is constructed for $J_2(\mathbf{A})$ to make the unmixing result more accurate.

A. Spectral Constraint for Endmembers

Minimum volume of simplex determined by endmembers has been demonstrated to be extremely effective to approximate an optimal solution of endmembers due to the concepts of convex [20]. However, when it is utilized to constrain an NMF based algorithm, dimension reduction is required as a preprocessing step. In addition, matrix determinant and matrix inversion computation may cause numerical instability when the simplex volume becomes small [23]. Actually, to find a simplex with minimum volume is in essence to find a simplex which is as compact as possible. Therefore, a minimum spectral distance constraint is alternatively imposed to enforce compactness of endmembers, which is defined as

$$J_1(\mathbf{M}) = \sum_{i=1}^{c} (\mathbf{m}_i - \bar{\mathbf{m}})^T (\mathbf{m}_i - \bar{\mathbf{m}}) \tag{14}$$

where $\mathbf{m}_i (i = 1, 2, \ldots, c)$ represents an endmember in $\mathbf{M}$, and $\bar{\mathbf{m}}$ defines the average spectrum of endmember matrix $\mathbf{M}$. In matrix form, this constraint can be expressed as follows:

$$J_1(\mathbf{M}) = \text{trace}\left((\mathbf{M}-\bar{\mathbf{M}})^T(\mathbf{M}-\bar{\mathbf{M}})\right). \tag{15}$$

The proposed minimum spectral distance constraint for endmembers, which utilizes cumulative distance between each endmember spectrum and the mean spectral signature to replace volume of the simplex formed by endmembers, can also optimize endmember spectra as compact as possible. Actually, in a two-dimensional (2-D) space, the simplex formed by endmembers is reduced to a triangle. The proposed endmemberwise distance measures the distance from the three vertices of the triangle to its center of gravity while the volume measures the area of the triangle. Both measures can ensure compactness of endmembers. However, the proposed spectral distance outperforms the volume of simplex for several reasons. First, it is a convex constraint by which the optimization algorithm for SMA will benefit a lot. Second, it does not need a dimension reduction preprocessing step. Finally, numerical instability, which is caused by matrix determinant and inversion operations when the volume of simplex becomes small, can be avoided.

B. Spatial Constraint for Abundance

Recent studies have shown that many real world data are actually sampled from a nonlinear low-dimensional manifold, which is embedded in the high-dimensional ambient space. In this article, the local geometric structure of data is exploited by considering the relationship between pixels and their corresponding neighbors. Therefore, the local linear embedding (LLE) assumption, which means that each pixel can be linearly reconstructed by its neighbors, is considered.

For each hyperspectral pixel $\mathbf{r}_i (i = 1, 2, \ldots, o)$ in the image, we use $N(\mathbf{r}_i)$ represents its neighboring pixels. According to the LLE assumption, $\mathbf{r}_i$ can be represented as the linear combination of its neighboring pixels:

$$\mathbf{r}_i = \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{r}_j \tag{16}$$

where W_{ij} represents the linear coefficient of $\mathbf{r}_j$ in the reconstruction of $\mathbf{r}_i$. Generally, the linear coefficients W_{ij} can be obtained by the following objective function:

$$W_{ij} = \arg\min \left\| \mathbf{r}_i - \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{r}_j \right\|_2. \tag{17}$$

According to the LMM [defined by (1)] and the LLE [defined by (16)], each pixel can be represented as follows:

$$\begin{aligned} \mathbf{r}_i &\approx \mathbf{M}\mathbf{a}_i \\ &\approx \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{r}_j \\ &\approx \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}(\mathbf{M}\mathbf{a}_j) \\ &= \mathbf{M} \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{a}_j. \end{aligned} \tag{18}$$

Observed from (18), we have

$$\mathbf{a}_i = \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{a}_j \tag{19}$$

which implies that the local geometric structure of data can be preserved by abundances in the unmixing procedure. Therefore, the following spatial preserving constraint can be imposed on abundances to preserve the relationship of spatial neighborhood:

$$\begin{aligned} J_2(\mathbf{A}) &= \sum_{i=1}^{o} ||\mathbf{a}_i - \sum_{\mathbf{r}_j \in N(\mathbf{r}_i)} W_{ij}\mathbf{a}_j||_2^2 \\ &= ||\mathbf{A} - \mathbf{A}\mathbf{W}||_F^2 \end{aligned}$$

$$= ||\mathbf{A}(\mathbf{I}-\mathbf{W})||_F^2$$
$$= \text{trace}(\mathbf{A}(\mathbf{I}-\mathbf{W})(\mathbf{I}-\mathbf{W})^T\mathbf{A}^T)$$
$$= \text{trace}(\mathbf{APP}^T\mathbf{A}^T) \quad (20)$$

where $trace(\cdot)$ denotes the operation of trace, $\mathbf{I} = \text{diag}(1, 1, \ldots, 1)$, and $\mathbf{P} = \mathbf{I} - \mathbf{W}$.

C. Spectral and Spatial Constrained NMF Algorithm for SMA

According to the spectral distance constraint for endmembers [defined by (15)] and the spatial preserving constraint for abundance (20), the proposed spectral and spatial constrained NMF (SS-NMF) model can be constructed as

$$\mathbf{F}(\mathbf{M},\mathbf{A}) = \frac{1}{2}\|\mathbf{R}-\mathbf{MA}\|_{2,1} + \frac{1}{2}\lambda_1 \cdot \text{trace}\left((\mathbf{M}-\bar{\mathbf{M}})^T(\mathbf{M}-\bar{\mathbf{M}})\right) + \frac{1}{2}\lambda_2 \cdot \text{trace}(\mathbf{APP}^T\mathbf{A}^T). \quad (21)$$

According to the PG method defined by Algorithm 1, the proposed SS-NMF model can be effectively solved by the following two suboptimization problems with iterative updates:

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}\succeq 0} \mathbf{F}(\mathbf{M}^k, \mathbf{A}^k)$$
$$\mathbf{M}^{k+1} = \arg\min_{\mathbf{M}\succeq 0} \mathbf{F}(\mathbf{M}^k, \mathbf{A}^{k+1}). \quad (22)$$

In other words, alternatively fixes one matrix and improves the other one. The convergence of this alternating method for NMF has been explained by Lin [59] and it has been pointed out that any limit point of the sequence $(\mathbf{M}^k, \mathbf{A}^k)$ generated by this method is a stationary point for NMF. Since both spectral distance constraint and spatial preserving constraint are second-order convex constraints, this conclusion is still held for the proposed SS-NMF algorithm.

Each suboptimization algorithm is a convex problem and can be solved by the PG method. Therefore, the gradient derivation of the proposed spectral distance constraint and spatial preserving constraint must be determined previously. The gradient derivation of the proposed spectral distance constraint is determined as

$$\nabla_{\mathbf{M}} J_1(\mathbf{M}) = 2 \cdot (\mathbf{M}-\bar{\mathbf{M}}) \cdot (\mathbf{I} - \frac{1}{c}\mathbf{1}) \quad (23)$$

where $\mathbf{I} = \text{diag}(1, 1, \ldots, 1)$, and $\mathbf{1}$ is a $\boldsymbol{c}$-by-$\boldsymbol{c}$ matrix of ones. The gradient derivation of the proposed spatial preserving constraint is determined as

$$\nabla_{\mathbf{A}} J_2(\mathbf{A}) = 2 \cdot \mathbf{APP}^T. \quad (24)$$

As a result, the gradient derivation of $\mathbf{F}(\mathbf{M},\mathbf{A})$ for these two subproblems of the proposed SS-NMF can be derived as

$$\nabla_{\mathbf{M}} f(\mathbf{M},\mathbf{A}) = \mathbf{Q}(\mathbf{MA}-\mathbf{R})\mathbf{A}^T + \lambda_1 \cdot (\mathbf{M}-\bar{\mathbf{M}}) \cdot \left(\mathbf{I} - \frac{1}{c}\mathbf{1}\right)$$
$$\nabla_{\mathbf{A}} f(\mathbf{M},\mathbf{A}) = \mathbf{M}^T\mathbf{Q}(\mathbf{MA}-\mathbf{R}) + \lambda_2 \cdot \mathbf{APP}^T. \quad (25)$$

Similar to the PG method of NMF, the nonnegative constraint can be guaranteed by a projective operation. The update rule for the proposed SS-NMF can be summarized as

$$\mathbf{A}^{k+1} = \max\left(0, \quad \mathbf{A}^k + \alpha_k \nabla_{\mathbf{A}} \mathbf{F}(\mathbf{M}^k, \mathbf{A}^k)\right)$$
$$\mathbf{M}^{k+1} = \max\left(0, \quad \mathbf{M}^k + \beta_k \nabla_{\mathbf{M}} \mathbf{F}(\mathbf{M}^k, \mathbf{A}^{k+1})\right) \quad (26)$$

where α_k and β_k are the stepsize. It has been pointed out that the stepsize is a critical value in the PG method. In this article, the well-known Armijo's technique is adopted to find the best stepsize for these two subproblems.

In addition, it should be noted that, in the optimization process of the proposed SS-NMF for SMA, the iteration stops when the maximum iteration number (set to 500) is reached or the following criterion is satisfied:

$$\frac{\mathbf{F}(\mathbf{M}^k,\mathbf{A}^k) - \mathbf{F}(\mathbf{M}^{k-1},\mathbf{A}^{k-1})}{\mathbf{F}(\mathbf{M}^{k-1},\mathbf{A}^{k-1})} < 10^{-4}. \quad (27)$$

IV. Experiments

In this section, extensive experiments on both synthetic dataset and real hyperspectral remote sensing image are conducted to validate the performance of the proposed SS-NMF algorithm.

A. Experiments Over Synthetic Dataset

A 100×100-pixel synthetic hyperspectral scene, which is simulated under linear mixture assumption using five spectra of minerals from a spectral library compiled by the U.S. Geological Survey (USGS)[1] containing as many as 420 bands covering from 400 to 2500 nm, is adopted in this experiment. The simulation strategy in [60] is adopted, which mainly involves two steps: pure-pixel image simulation and mixed-pixel image simulation. The whole image is divided into 25 square regions with the same size of 20×20 pixels. All the pixels in the same small region are assigned as a homogeneous ground object. In each row, five ground objects are assigned randomly to the five regions. Furthermore, 10-dB Gaussian noise is added to the whole image and 50-dB Gaussian noise is added to the pixels inside a block to simulate spectral variability. In order to generate mixed pixels in the image, a 15×15 window is used for local averaging via spatial convolution. Fig. 1 shows the five reflectance curves of selected spectra from the USGS spectral library.

To assess the performance of the proposed algorithm, the spectral angle distance (SAD) is adopted to evaluate the performance of estimated endmember in SMA. Let $\hat{\mathbf{m}}_i$ be an estimated endmember and $\mathbf{m}_i$ be the most similar spectral signature in the USGS library. The SAD value between two spectral signatures is defined as

$$\text{SAD}(\mathbf{m}_i, \hat{\mathbf{m}}_i) = \arccos\left(\frac{\mathbf{m}_i \cdot \hat{\mathbf{m}}_i}{\|\mathbf{m}_i\| \cdot \|\hat{\mathbf{m}}_i\|}\right) \quad (28)$$

where $\|\cdot\|$ represents the magnitude of vectors. Note that lower SAD scores mean higher spectral similarity between the compared vectors. In the synthetic experiment, abundance

[1][Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

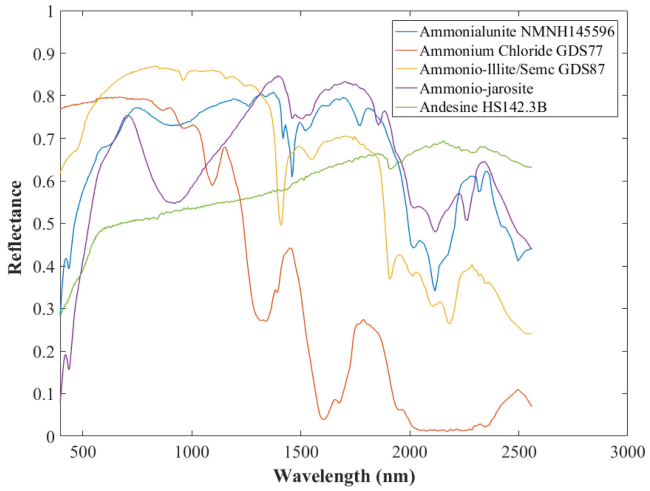


Fig. 1. Five typical hyperspectral signatures of typical minerals from the USGS digital spectral library.

fractions of each endmember are also known as ground-truth. Therefore, the abundance estimation error (labeled as "EstEr"), which is evaluated by the root-mean-square error between the true abundance and the estimated fractional abundance, is also evaluated for quantitative evaluation. The lower the EstEr, the better the abundance reconstruction performance. Let $\hat{\mathbf{a}}_i$ denote the SS-NMF estimated abundance of $\mathbf{m}_i$, with $\mathbf{a}_i$ being the corresponding ground-truth abundance. Therefore, the EstEr is defined as

$$\mathrm{EstEr}(\mathbf{a}_i, \hat{\mathbf{a}}_i) = \frac{1}{o}\left(\sum_{j=1}^{o}(\boldsymbol{a}_{ij} - \hat{\boldsymbol{a}}_{ij})^2\right)^{\frac{1}{2}}. \tag{29}$$

Moreover, the signal-to-reconstruction error (SRE) is also adopted to measure the quality of the reconstruction of whole image, which is defined as

$$\mathrm{SRE}(\mathbf{R}, \mathbf{M}, \mathbf{A}) = 10\mathrm{lg}\frac{\mathrm{E}[\|\mathbf{R}\|_2^2]}{\mathrm{E}[\|\mathbf{R} - \mathbf{MA}\|_2^2]}. \tag{30}$$

It measures the information regarding the power of the error in relation to the power of the signal. Generally, the higher the SRE (dB), the better the image reconstruction performance.

1) Parameter Analysis: The proposed SS-NMF algorithm involves two parameters to balance the importance of spectral constraint and spatial constraint, i.e., λ_1 and λ_2 in (13). The unmixing results of the proposed SS-NMF algorithm with different values of λ_1 and λ_2 are shown in Fig. 2, where λ_1 varies exponentially from $1\mathrm{e}-7$ to 1e2 and λ_2 varies exponentially from $1\mathrm{e}-3$ to 1e6. To balance the importance of different constraints, the optimal combination of λ_1 and λ_2 is set to $(1\mathrm{e}-5, 1)$.

2) Ablation Experiment: With respect to the proposed method, it mainly consists of three parts: $L_{2,1}$-norm based loss function, spectral endmember constraint, and spatial abundance constraint. To verify the efficiency of each part, we separate those parts and conduct an ablation experiment. As shown in Table I, four other variants of the proposed SS-NMF are considered:

TABLE I
SPECIFIC EXPERIMENTAL CONFIGURATION OF THE PROPOSED SS-NMF VARIANTS

Variants	Loss Function	Spectral Constraint $J_1(\mathbf{M})$	Spatial Constraint $J_2(\mathbf{A})$
SS-NMF$_F$	Frobenius	√	√
NMF$_{2,1}$	$L_{2,1}$	×	×
S$_M$-NMF	$L_{2,1}$	√	×
S$_A$-NMF	$L_{2,1}$	×	√
SS-NMF	$L_{2,1}$	√	√

TABLE II
AVERAGE SAD SCORES (IN DEGREES), ESTER VALUE, AND SRE VALUE OF THE SS-NMF VARIANTS ON THE SYNTHETIC DATASET

Algorithms	SS-NMF$_F$	NMF$_{2,1}$	S$_M$-NMF	S$_A$-NMF	SS-NMF
SAD (endmember)	4.1376	0.2782	0.2790	0.2780	**0.2706**
EstEr (abundance)	0.2931	0.2627	0.2482	0.2608	**0.2304**
SRE (image)	24.2659	23.0081	23.0069	23.1217	**25.5317**

The best results are in bold.

1) SS-NMF$_F$: Frobenius-norm based loss function combining with both spectral endmember constraint and spatial abundance constraint;
2) NMF$_{2,1}$: only $L_{2,1}$-norm based loss function;
3) S$_M$-NMF: $L_{2,1}$-norm based loss function combining with spectral end member constraint;
4) S$_A$-NMF: $L_{2,1}$-norm based loss function combining with spatial abundance constraint.

The synthetic experimental results of all these variants are listed in Table II. It is observed that

1) the $L_{2,1}$-norm based loss function is better to model the reconstruction error than L_F-norm when NMF-based unmixing model is used;
2) both the proposed spatial neighborhood preserving constraint and the proposed minimum spectral distance constraint can certainly improve the performance of $L_{2,1}$-norm based NMF model for SMA.
3) when the proposed spatial neighborhood preserving constraint and the proposed minimum spectral distance constraint are jointly imposed over $L_{2,1}$-norm-based NMF model, the performance of SMA can be further improved, demonstrating the necessity and complementary of these two spectral and spatial constraints for NMF model.

3) Comparison With Typical NMF-Based Algorithms: Several well-known NMF-based algorithms are adopted for comparison, including MVC-NMF [20], NMF-QMV (TV) [22], and SGSNMF [35], together with the baseline comparison of VCA [61]+FCLS [62]. In all cases, parameters of these comparison methods have been carefully optimized so that the best performance for each method is reported. To ensure reliable comparisons, the VCA and FCLS algorithms are adopted to initialize all the endmember matrixes and abundance matrixes. The computational time of comparison algorithms is also measured with Inter Core i7-9700 CPU at 3000 GHz with 16 GB of RAM using MATLAB R2020b on Windows 10 platform. Moreover, 20 independent runs have been carried out and the results are averaged for the proposed and comparison algorithms in the experiments.

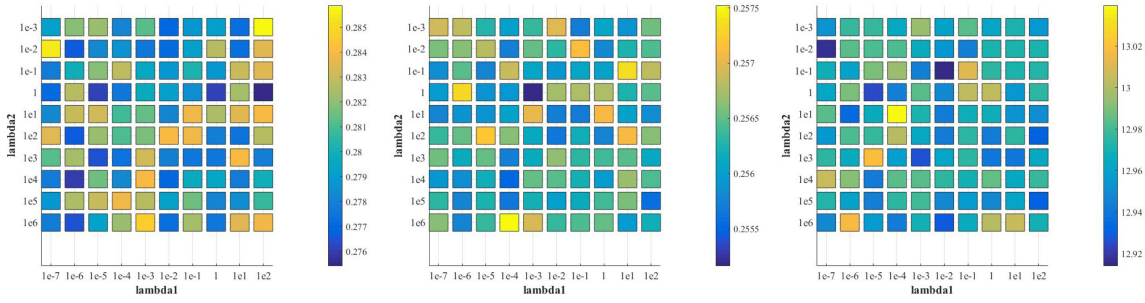


Fig. 2. Parameter analysis for the synthetic dataset, evaluated by SAD for endmembers, EstEr for abundance, and SRE for pixels.

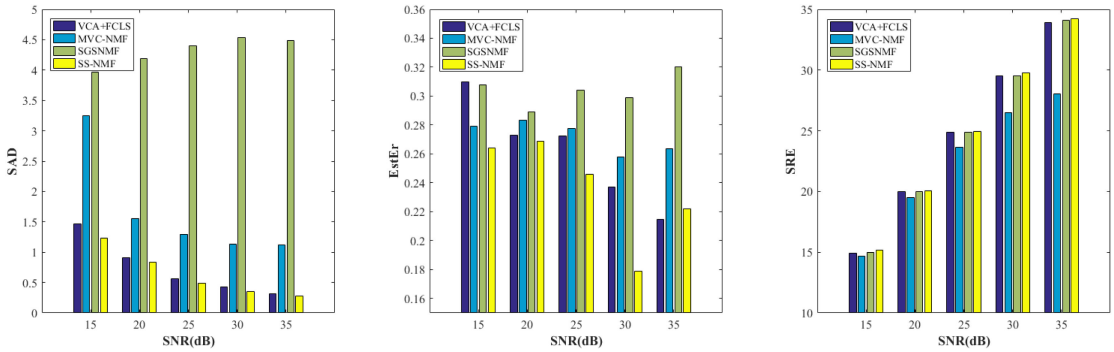


Fig. 3. Noise sensitivity analysis of different algorithms over the synthetic image under different SNRs, evaluated by SAD for endmembers, EstEr for abundance, and SRE for image.

TABLE III
AVERAGE SAD SCORES (IN DEGREES), ESTER VALUE, SRE VALUE, AND COMPUTATIONAL TIME (IN SECONDS) OF COMPARISON ALGORITHMS ON THE SYNTHETIC DATASET

Algorithms	VCA+FCLS	MVC-NMF	SGSNMF	NMF-QMV	**SS-NMF**
SAD (endmember)	0.4215	1.0865	0.2993	0.3280	**0.2706**
EstEr (abundance)	0.2727	0.2573	0.2401	0.2657	**0.2304**
SRE (image)	14.0074	14.0561	20.0851	20.5119	**25.5317**
Time (process)	**1.6263**	381.4362	34.5617	13.5120	48.9941

The best results are in bold.

Table III shows the average SAD and EstEr scores between the ground-truth endmembers or abundance and corresponding matrixes acquired by different algorithms over synthetic dataset, as well as SRE value of the whole image and computational time of each algorithm operation. It is observed that the proposed SS-NMF method obviously outperforms other considered algorithms by achieving the lowest SAD and EstEr values and the highest SRE value. However, due to the double constraints of proposed SS-NMF, it needs more time to find optimal solutions than that of NMF-QMV using spatial constraints.

In order to analyze the sensitivity of the proposed SS-NMF to noises, different noise levels involved in the synthetic dataset are also considered by varying SNR from 15 to 35 dB with the interval of 5 dB. The unmixing performance for all these algorithms over the synthetic dataset with different SNRs is shown in Fig. 3. It is observed that the proposed SS-NMF outperforms under different noise levels. Moreover, the unmixing performance of the proposed SS-NMF steadily increased in terms of SAD and SRE when the SNR increases.

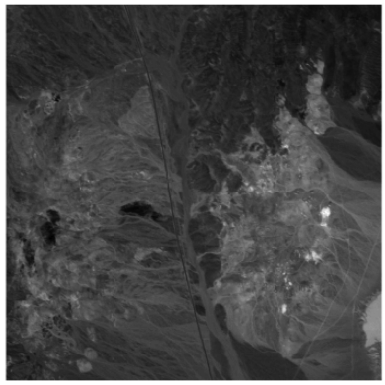
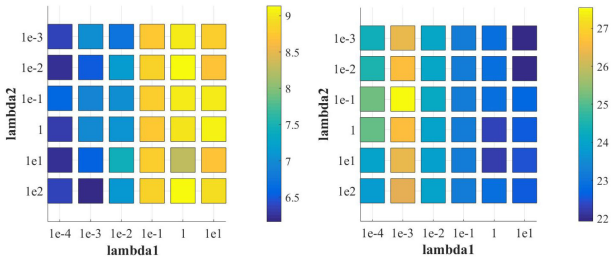


Fig. 4. Band 100 of the subimage of AVIRIS Cuprite dataset.

B. Experiments Over Real Dataset

The well-known Cuprite dataset acquired by AVIRIS[2] contains 224 bands ranging from 370 to 2510 nm with a ground instantaneous field of view of 20 m. A 350×350-pixel subset from the sector labeled as $f970619t01p02_r02_sc03.a.rfl$, which is shown in Fig. 4, is used for performance validation and the number of endmembers are set to 16. After removing noisy bands

[2][Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html



(a) (b)

Fig. 5. Parameter analysis for the Cuprite dataset, evaluated by SAD for endmembers and SRE for image.

TABLE IV
AVERAGE SAD SCORES (IN DEGREES) AND SRE VALUE OBTAINED BY THE SS-NMF VARIANTS ON THE AVIRIS CUPRITE DATASET

Algorithms	SS-NMF$_F$	NMF$_{2,1}$	S$_M$-NMF	S$_A$-NMF	SS-NMF
Alunite	**5.8032**	10.3481	11.7379	11.2778	7.2405
Buddingtonite	6.4346	**5.7923**	6.2513	5.8279	7.1259
Kaolinite	9.3814	7.3938	9.3393	7.0243	**6.7619**
Montmorillonite	9.7735	4.5493	**3.9558**	4.6193	5.5660
Muscovite	6.6410	7.3312	6.6603	7.3502	**6.2480**
Jarosite	7.5315	7.3998	7.5095	7.3513	**6.8178**
Andradite	6.2606	6.1751	6.2443	6.1908	**4.4376**
Dumortierite	6.4924	5.2086	5.8154	5.0701	**4.6408**
Mean SAD	7.2898	6.7748	7.1892	6.8390	**6.1048**
SRE (dB)	17.7283	23.2731	21.7646	22.7932	**28.3149**

Note that the best results of each signature and mean values are in bold.

and water-absorption bands (including bands $1-4$, $105-115$, $150-170$, and $223-224$), a total of 186 reflectance bands are finally adopted. In this experiment, for quantitative evaluation, only eight reference signatures are selected from the USGS digital spectral library, corresponding to highly representative minerals in the Cuprite mining district, including Alunite, Andradite, Buddingtonite, Dumortierite, Jarosite, Kaolinite, Montmorillonite, and Muscovite. For quantitatively evaluating the performance of our proposed SS-NMF algorithm over Cuprite dataset, two criteria, SAD for endmember and SRE for image as above, are adopted.

1) Parameter Analysis: The unmixing results of the proposed SS-NMF algorithm under the two parameters λ_1 and λ_2 are shown in Fig. 5, where λ_1 was set from $1\mathrm{e}-4$ to 1e1, λ_2 from $1\mathrm{e}-3$ to 1e2 exponentially. To balance the performance of estimated endmember and reconstructed pixels, the optimal combination of λ_1 and λ_2 is $(1\mathrm{e}-3, 1\mathrm{e}2)$.

2) Ablation Experiment: The ablation experiment using different variants of the proposed SS-NMF algorithm is also considered over Cuprite dataset, and the results are shown in Table IV. It is also confirmed that the proposed spatial neighborhood preserving constraint and the proposed minimum spectral distance constraint are very effective to constrain NMF model for SMA, and they can jointly perform the best for unmixing. Moreover, the $L_{2,1}$-norm based loss function is more efficient to model noise and outliers than L_F-norm in the real scene.

3) Comparison With Typical NMF-Based Algorithms: The three typical NMF-based algorithms, i.e., MVC-NMF [20], NMF-QMV (TV) [22], and SGSNMF [35], together with the baseline of VCA [61]+FCLS [62], are also adopted for comparison in this experiment. The experimental results of all these algorithms over the AVIRIS Cuprite dataset are listed in Table V, which lists the SAD scores between each acquired endmember and corresponding selected USGS spectral signature, as well as the mean SAD values across all eight selected signatures and the SRE value of the entire image. It is observed that all the algorithms obtain better spectral approximations for certain different minerals, while those by the proposed SS-NMF can reach the lowest mean SAD value and highest SRE value, indicating that the estimated endmembers and reconstructed pixels by SS-NMF generally provide a good match with regard to the corresponding reference signatures as well as the original pixels. Meanwhile, the computational time of SS-NMF is less than that of MVC-NMF and NMF-QMV, which verifies the high efficiency of proposed model in complex spatial context and under larger number of pixels.

TABLE V
AVERAGE SAD SCORES (IN DEGREES), SRE VALUE, AND COMPUTATIONAL TIME OBTAINED BY COMPARISON ALGORITHMS ON THE AVIRIS CUPRITE DATASET

Algorithms	VCA+FCLS	MVC-NMF	SGSNMF	NMF-QMV	**SS-NMF**
Alunite	9.4545	8.2778	7.7864	**5.0787**	7.2405
Buddingtonite	**4.0214**	5.8279	4.9542	8.2191	7.1259
Kaolinite	6.4516	7.0243	5.5007	**4.6754**	6.7619
Montmorillonite	6.1501	4.6193	**3.9781**	7.9922	5.5660
Muscovite	7.0865	6.5502	6.9464	6.2842	**6.2480**
Jarosite	7.4594	7.3513	**6.1528**	7.8170	6.8178
Andradite	6.2984	5.8908	7.0034	5.6352	**4.4376**
Dumortierite	**4.2832**	4.9701	7.3011	5.1186	4.6408
Mean SAD	6.4006	6.3140	6.2029	6.3526	**6.1048**
SRE (dB)	12.9450	16.2950	20.1041	12.8676	**28.3149**
Time (second)	**90.4773**	4717.5650	142.9965	860.2000	673.3352

Note that the best results of each signature and mean values are in bold.

Fig. 6 shows a qualitative comparison between the abundance maps of three typical signatures: 1) Alunite+Muscovite and/or Kaolinite; 2) Buddingtonite; and 3) Montmorillonite, with the corresponding reference classification maps produced by Tricorder software (labeled as "GT"). It can be seen that the proposed SS-NMF algorithm produces clearest abundance maps in close agreement with the ground truth of the mineral maps, which is most consistent and stable. To sum up, these experimental results confirm that the proposed SS-NMF algorithm generates the best unmixing performance.

C. Convergence Analysis

Both synthetic and Cuprite datasets are used to analyze the convergence of the proposed method. The convergence curves of SS-NMF on two data are presented in Fig. 7, where we show how the objective function value $\mathbf{F}(\mathbf{M}, \mathbf{A})$ changes as it evolves over 1000 iterations. As shown in Fig. 7, the proposed algorithm converges well on both synthetic and real data. SS-NMF on synthetic data converges in about 200 iterations while Cuprite data converges in about 400 iterations. Therefore, we set a maximum number of iterations of 500 is reasonable in this article, and this experiment can verify that the proposed SS-NMF algorithm exhibited a good convergence behavior.

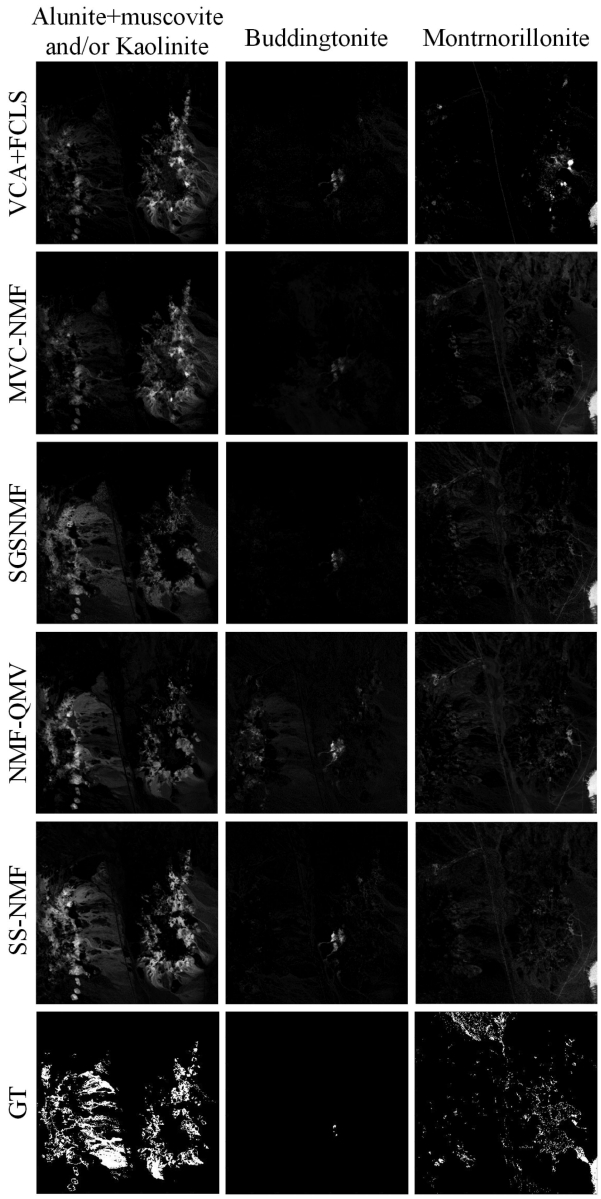
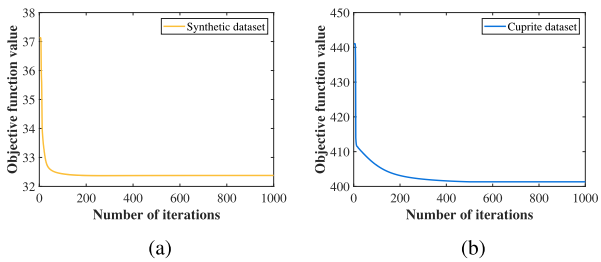


Fig. 6. Abundance maps obtained by different algorithms over Cuprite dataset for three typical spectral signatures. (Black: 0% abundance and White: 100% abundance).



(a) (b)

Fig. 7. Convergence analysis of SS-NMF on (a) Synthetic dataset and (b) Cuprite dataset.

V. Conclusion

In this article, a spectral-spatial constrained NMF (SS-NMF) is proposed for SMA of hyperspectral images by imposing a spectral constraint over endmembers and a spatial constraint over abundance simultaneously in the NMF-based unmixing model. Specifically, an $L_{2,1}$-norm loss function is employed to effectively handle non-Gaussian noises or outliers. Meanwhile, a minimum spectral distance constraint that optimizes endmember spectra as compact as possible and a spatial preserving constraint to preserve the local structure of the data on a low-dimensional manifold are imposed to enhance SMA. Experiments over both synthetic dataset and well-known Cuprite dataset are conducted to verify the effectiveness and superiority of the proposed algorithm over several typical NMF-based algorithms. The experimental results show that the proposed SS-NMF algorithm outperforms the other NMF-based algorithms both qualitatively and quantitatively, which further reveals that the proposed spatial and spectral constraints have a significant improvement on the unmixing results. Although current performance is very encouraging, the long computational time of the solution method is still a challenge for practical applications. A more efficient optimization algorithm will be studied in our future research.

References

[1] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 629-647, Jan. 2021.

[2] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[3] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[4] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon, "Matrix cofactorization for joint spatial-spectral unmixing of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4915–4927, Jul. 2020.

[5] R. Tao, X. Zhao, W. Li, H. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[6] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 100–111, Jan. 2020.

[7] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.

[8] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.

[9] X. Zhao *et al.*, "Joint classification of hyperspectral and lidar data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[10] D. Heinz and C. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[11] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1008–1021, Mar. 2007.

[12] 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.

[13] C.-I. Chang, S.-S. Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 375–392, Feb. 2002.

[14] J. M. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 4033–4036.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[16] X. Tao, B. Wang, L. Zhang, and J. Zhang, "A new scheme for decomposition of mixed pixels based on nonnegative matrix factorization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 1759–1762.

[17] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[18] V. Pauca, J. Piper, and R. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.

[19] S. Mei, M. He, Y. Zhang, Z. Wang, and D. Feng, "Improving spatial-spectral endmember extraction in the presence of anomalous ground objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4210–4222, Nov. 2011.

[20] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, May 2007.

[21] A. M. S. Ang and N. Gillis, "Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4843–4853, Dec. 2019.

[22] L. Zhuang, C. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9858–9877, Dec. 2019.

[23] Y. Yu, S. Guo, and W. Sun, "Minimum distance constrained non-negative matrix factorization for the endmember extraction of hyperspectral images," *Proc. SPIE*, vol. 6790, 2007, pp. 1–9.

[24] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.

[25] L. Tong, J. Zhou, Y. Qian, X. Bai, and Y. Gao, "Nonnegative matrix factorization based hyperspectral unmixing with partially known endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 1–14, Nov. 2016.

[26] 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.

[27] F. Hao, A. Li, H. Xu, and W. Tao, "Sparsity-constrained deep nonnegative matrix factorization for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1105–1109, Jul. 2018.

[28] W. Wang, Y. Qian, and Y. Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 681–694, Feb. 2016.

[29] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[30] X. Feng, H. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.

[31] R. Liu, D. Bo, and Z. Liangpei, "Hyperspectral unmixing via double abundance characteristics constraints based NMF," *Remote Sens.*, vol. 8, no. 6, pp. 464–486, 2016.

[32] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.

[33] S. Yang, X. Zhang, Y. Yao, S. Cheng, and L. Jiao, "Geometric nonnegative matrix factorization (GNMF) for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2696–2703, Jun. 2015.

[34] L. Tong, J. Zhou, X. Li, Y. Qian, and Y. Gao, "Region-based structure preserving nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1575–1588, Apr. 2017.

[35] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017.

[36] L. Yang, J. Peng, H. Su, L. Xu, Y. Wang, and B. Yu, "Combined nonlocal spatial information and spatial group sparsity in NMF for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1767–1771, Oct. 2020.

[37] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[38] Y. E. Salehani and S. Gazor, "Smooth and sparse regularization for NMF hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3677–3692, Aug. 2017.

[39] H. Fang, A. Li, T. Wang, and H. Xu, "Hyperspectral unmixing using double-constrained multilayer NMF," *Remote Sens. Lett.*, vol. 10, no. 3, pp. 224–233, 2019.

[40] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.

[41] Y. Yuan, Z. Zhang, and Q. Wang, "Improved collaborative non-negative matrix factorization and total variation for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 998–1010, Mar. 2020.

[42] K. Qu and W. Bao, "Multiple-priors ensemble constrained nonnegative matrix factorization for spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 963–975, Feb. 2020.

[43] L. Tong, J. Zhou, B. Qian, J. Yu, and C. Xiao, "Adaptive graph regularized multilayer nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 434–447, Jan. 2020.

[44] J. Peng, Y. Zhou, W. Sun, Q. Du, and L. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1501–1515, Feb. 2021.

[45] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, Apr. 2019.

[46] H. C. Li, S. Liu, X. R. Feng, and S. Q. Zhang, "Sparsity-constrained coupled nonnegative matrix-tensor factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5061–5073, Aug. 2020.

[47] X. R. Feng, H. C. Li, S. Liu, and H. Zhang, "Correntropy-based autoencoder-like NMF with total variation for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2020.3020896.

[48] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, and C. Richard, "Deep generative models for library augmentation in multiple endmember spectral mixture analysis," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2020.3007161.

[49] Z. Han, D. Hong, L. Gao, B. Zhang, and J. Chanussot, "Deep half-siamese networks for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2020.3011941.

[50] S. Mei, M. He, Z. Wang, and D. Feng, "Mixture analysis by multichanel hopfield neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 455–459, Jul. 2010.

[51] M. Yong, L. Chang, X. Mei, C. Liu, and J. Ma, "Robust sparse hyperspectral unmixing with $L_{2,1}$ norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, Mar. 2017.

[52] J. Li, Y. Li, R. Song, S. Mei, and Q. Du, "Local spectral similarity preserving regularized robust sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7756–7769, Oct. 2019.

[53] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint l2,1-norms minimization," *Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1813–1821.

[54] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Proc. 13th Int. Conf. Adv. Neural Inf. Process. Syst.*, 2001, pp. 535–541.

[55] M. Chu, F. Diele, R. Plemmons, and S. Ragni, "Optimality, computation, and interpretation of nonnegative matrix factorizations," *Tech. Rep.*, NCSU, 2005

[56] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 713–730, 2008.

[57] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM J. Sci. Comput.*, vol. 33, no. 6, pp. 3261–3281, 2011.
[58] D. Bertsekas, "On the Goldstein–Levitin–Polyak gradient projection method," *IEEE Trans. Autom. Control*, vol. 21, no. 2, pp. 174–184, 2003.
[59] C. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, 2007.
[60] S. Mei, M. He, Z. Wang, and D. Feng, "Spatial purity based endmember extraction for spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3434–3445, Sep. 2010.
[61] J. Nascimento and J. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[62] D. Heinz, C.-I. Chang, and M. L. G. Althouse, "Fully constrained least squares-based linear unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 1999, pp. 1401–1403.

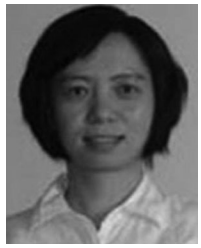
Ge Zhang (Student Member, IEEE) received the B.S. degree in electronics and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2018. She is currently working toward the Ph.D. degree in information and communication engineering with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.

Her research interests include hyperspectral image unmixing and signal processing.

Shaohui Mei (Senior Member, IEEE) received the B.S. degree in electronics and information engineering and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2005 and 2011, respectively.

He is currently an Associate Professor with the School of Electronics and Information. He was a Visiting Student with the University of Sydney, from October 2007 to October 2008. His research interests include hyperspectral remote sensing image processing and applications, intelligent signal and information acquisition and processing, video processing, and pattern recognition.

Dr. Mei was the recipient of Excellent Doctoral Dissertation Award of Shaanxi Province in 2014, and Best Paper Award of IEEE ISPACS 2017. He is currently serving as Associated Editor of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the Reviewer of more than 20 international famous academic journals, and was awarded the Best Reviewer of IEEE JSTARS in 2019. He served as the Registration Chair of IEEE China Summit and International Conference on Signal and Information Processing 2014.



Yan Feng received the B.S. degree in electronic engineering and the M.S. and Ph.D. degrees in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 1984, 1989, and 2007, respectively.

She is currently a Professor with the School of Electronics and Information, Northwestern Polytechnical University. Her research interests include image processing, compression and classification of hyperspectral images, compressed sensing theory, and intelligent information processing.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore County, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIE-International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was a Co-Chair for the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair for Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She served as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She is the General Chair for the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing in Shanghai, China, in 2012.