





Nonconvex Nonseparable Sparse Nonnegative Matrix Factorization for Hyperspectral Unmixing

Fengchao Xiong , *Member, IEEE*, Jun Zhou , *Senior Member, IEEE*, Jianfeng Lu , *Member, IEEE*,
and Yuntao Qian , *Member, IEEE*

Abstract—Hyperspectral unmixing is an important step to learn the material categories and corresponding distributions in a scene. Over the past decade, nonnegative matrix factorization (NMF) has been utilized for this task, thanks to its good physical interpretation. The solution space of NMF is very huge due to its nonconvex objective function for both variables simultaneously. Many convex and nonconvex sparse regularizations are embedded into NMF to limit the number of trivial solutions. Unfortunately, they either produce biased sparse solutions or unbiased sparse solutions with the sacrifice of the convex objective function of NMF with respect to individual variable. In this article, we enhance NMF by introducing a generalized minimax concave (GMC) sparse regularization. The GMC regularization is nonconvex and nonseparable, enabling promotion of unbiased and sparser results while simultaneously preserving the convexity of NMF for each variable separately. Therefore, GMC-NMF better avoids being trapped into local minimals, and thereby produce physically meaningful and accurate results. Extensive experimental results on synthetic data and real-world data verify its utility when compared with several state-of-the-art approaches.

Index Terms—Generalized minimax concave (GMC) regularization, hyperspectral unmixing, nonnegative matrix factorization (NMF), sparse representation.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) simultaneously records 2-D spatial information and 1-D spectrum indexed by numerous narrow spectral bands of the target. The spectral bands significantly augment the discriminative ability, making HSI widely adopted in both remote sensing and computer vision tasks [1], [2]. However, due to the limited spatial resolution of hyperspectral sensors and long-distance imaging, many HSI pixels contain more than one type of materials, causing many

problems in practical applications. By decomposing a pixel into the combination of a bundle of constituted materials (endmembers) and corresponding fractions (abundances), hyperspectral unmixing provides an attractive way to tackle this problem, therefore benefiting subsequent applications.

Hyperspectral unmixing models can be broadly categorized into linear type and nonlinear type. Nonlinear spectral mixture models assume multiple-order physical interactions between lights scattered by more than one material. Representative models include bilinear models [3], multilinear models [4], and graphical models [5], just to name a few. Linear spectral mixture model (LMM) describes the light interacting with one material and linearly represents a pixel with endmembers. Under the framework of LMM, pure pixel-based methods such as pixel purity index [6] and vertex component analysis (VCA) [7] extract endmembers from the data by assuming the presence of pixels which include only one type of material. Subsequently, linear regression with nonnegativity and sum-to-one constraints can be adopted for abundance estimation [8]. The assumption of the pure pixels may not be realistic in complex scenarios, in which all the pixels are highly mixed with no pure pixels existing. Alternatively, the unmixing can be considered as a blind source separation problem, which simultaneously infers the endmembers and the abundances with no assumption of the presence of pure pixels. Nonnegative matrix factorization (NMF), which decomposes the data into the product of two matrices, stands out among these methods [9]–[11]. Thanks to its inherent advantages of learning parts-based representation of the data, the produced results are more physically interpretable and understandable, i.e., two matrices can be considered as endmember matrix and abundance matrix, respectively.

However, the solution space of NMF is very huge due to its nonconvex objective function. Many constraints are added to NMF to reduce the number of trivial solutions. So far, various constraints, such as minimum simplex volume [12] and endmember dissimilarity [13], were embedded into NMF for more physically meaningful endmember spectral signatures. Since hyperspectral images feature local/nonlocal similarity property, similar pixels are very likely to maintain this property in abundance maps. Following this property, some studies aim to integrate the spatial/spatial-spectral contextual information into NMF to guide accurate abundance estimation [14], [15]. Lu *et al.* [16] assumed hyperspectral data lie on a low-dimensional manifold subspace embedded in high-dimensional spectrum and

Manuscript received July 2, 2020; revised August 23, 2020 and September 9, 2020; accepted September 25, 2020. Date of publication October 1, 2020; date of current version October 15, 2020. This work was supported in part by the 2030 National Key AI Program of China under Grant 2018AAA0100500, in part by the National Key Research and Development Program of China under Grants 2017YFB1300205 and 2018YFB0505000, and in part by the National Natural Science Foundation of China under Grant 62002169. (*Corresponding author: Fengchao Xiong.*)

Fengchao Xiong and Jianfeng Lu are with the College of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: fcxiong@njust.edu.cn; luji@njust.edu.cn).

Jun Zhou is with the School of Information and Communication Technology, Griffith University, Nathan QLD 4111, Australia (e-mail: jun.zhou@griffith.edu.au).

Yuntao Qian is with the College of Computer Science, Zhejiang University, Hangzhou 310027, China (e-mail: ytqian@zju.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3028104

added manifold regularization to enforce local geometry structure information preserved in abundance maps. Yao *et al.* [17] encoded nonlocal total variation as spatial structural information explorer and added it to NMF to make nonlocal similar patches share smooth structure. Moreover, Peng *et al.* [10] employed self-paced NMF to selectively determine appropriate atoms, i.e., bands or pixels or elements for unmixing, so as to suppress the effect of noises.

Besides spatial information, sparsity is another widely recognized *prior* to narrow the solution space of NMF [18], [19]. Sparsity constraint assumes that a pixel is the mixture of only a small subset of endmembers, i.e., there are many zero entries in abundances. In the literature, most sparse regularizations belong to the class of additively separable functions, i.e., they take the form of $\Phi(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \phi_i(x_i)$, with $\phi_i: \mathbb{R} \rightarrow \mathbb{R}$ being the sparse-inducing regularization. Particularly, L_0 regularization directly minimizes the number of nonzero components and can yield a sparse solution, but leads to NP-hard optimization. To overcome the drawbacks of L_0 norm, it is a common practice to employ L_1 norm for sparse abundances because of its convex property, simplifying the optimization problem [20]. Though convex, L_1 norm-based sparse regularization tends to punish more on high-amplitude elements of variables which yet take up the majority of target signals, thereby producing biased estimation of abundances [21], [22]. Alternatively, improved sparse solutions are achieved by replacing L_1 regularization with nonconvex sparse regularizations [17], [18], [23]–[25], as they can better approximate L_0 norm and punish large amplitudes less than convex L_1 regularization does. For example, Qian *et al.* [26] added $L_{1/2}$ regularization to NMF ($L_{1/2}$ -NMF) for sparse unmixing and yielded more accurate abundance estimation. Taking advantages of an iterative half-thresholding algorithm, $L_{1/2}$ -NMF was further enhanced to overcome the complex nonconvex optimization problem of the original $L_{1/2}$ regularization [18]. Afterwards, Sigurdsson *et al.* [27] generalized $L_{1/2}$ sparse regularization to L_p ($0 < p < 1$) sparse regularization and obtained sparser and less-biased solutions in low signal-to-noise ratio (SNR) cases.

The major limitation of the above nonconvex separable sparse regularizations is that they induce sparsity by generally giving up the attractive convex characteristic of original NMF with respect to the individual variable. As a consequence, they introduce extraneous suboptimal local minimizers, making extended NMFs highly rely on the initializations and parameter settings [28]–[30]. Therefore, it is always expected that the sparse regularization can promote less-biased sparsity while simultaneously maintaining the benefits of the convex objective function. Some nonconvex nonseparable penalties can satisfy this requirement by coupling the sparse regularization with data fidelity term so that the objective function remains convex even though the penalty itself is nonconvex [31]. The nonconvexity leads to unbiased and sparser results and the nonseparability enables the regularization to consider the statistical relationships among the sparse signals [28]. Generalized minimax concave (GMC) penalty is such a nonconvex nonseparable sparse regularization with preserved convex objective function under mild condition [32]. Mathematically, GMC penalty is defined as a smooth

nonconvex function via infimal convolution subtracted from the L_1 norm. Compared with L_1 norm, it avoids systematic underestimation of high components of sparse vectors so that more accurate sparse approximation can be achieved. Unlike traditional nonconvex regularization, GMC penalty preserves the convex formulation of the sparsity-regularized objective function, making it share the attractive benefits of convex optimization such as the absence of spurious local minima. Thanks to these superior merits, GMC has been successfully applied in signal smoothing, bearing fault diagnosis, image registration, low-rank approximation, and more [22], [33]–[35]. To be specific, a convexity-edge-preserving smoother was developed in [22] by replacing L_1 norm-based total variation with GMC penalty, showing more favorable sharp edges in recovered signals. Brbic and Kopriva extended the GMC penalty to a low-rank sparse subspace clustering problem [21]. With the help of the GMC penalty, the rank and sparsity of data representation can be more accurately estimated than traditional nuclear and L_1 norms.

In this article, we introduce the GMC penalty to NMF, named GMC-NMF, to reduce the solution space of NMF-based unmixing. The advantages of GMC penalty in sparsity promoting and merits of convexity preserving make it more capable of deriving sparser abundances with less local minimal. Moreover, we derive the update rule under the framework of the multiplicative update and forward–backward splitting. Experimental results on synthetic data and real-world data show dominant advantages of the proposed GMC-NMF by comparing with alternative sparsity regularized NMFs.

The rest of this article is organized as follows. Section II introduces the background of the proposed method. In Section III, the details of the proposed GMC-NMF are presented and its properties are also highlighted. Section IV reports the experimental results on both synthetic and real-world data. The conclusion is drawn in Section V.

II. BACKGROUND

A. Linear Spectral Mixture Model

Based on the assumption that the incoming light undergoes interactions with a single material, LMM considers the observed spectrum as the linear combination of a set of endmembers. Given an HSI with L bands, LMM can be mathematically formulated as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^{L \times 1}$ is the spectral vector, $\mathbf{A} \in \mathbb{R}^{L \times R}$ is the end-member matrix containing R materials, $\mathbf{s} \in \mathbb{R}^{R \times 1}$ is the abundance vector corresponding to the contribution of each endmember, and $\mathbf{e} \in \mathbb{R}^{L \times 1}$ is an additive error vector accounting for the measurement errors. Under matrix notation, (1) is denoted by

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (2)$$

where \mathbf{X} is the HSI with L bands and N pixels, \mathbf{S} is an $R \times N$ abundance matrix whose each column is the fractional abundance for each pixel, and $\mathbf{E} \in \mathbb{R}^{L \times N}$ is an additional noise matrix.

Abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC) are usually added to LMM to satisfy the geometric interpretation that the acquired spectra are in the convex hull spanned by the endmembers. ANC indicates no endmember has negative fractions. ASC means the total contribution of all the endmembers should equal one.

B. Sparse Nonnegative Matrix Factorization

NMF seeks to approximate the original data matrix by decomposing it into the product of two low-dimensional nonnegative matrices. NMF provides a tractable model to automatically extract meaningful components and estimate associated proportions, which is respectively in line with endmembers and abundances in hyperspectral unmixing. Mathematically, NMF targets to solve the following objective function:

$$\min_{\mathbf{A}, \mathbf{S}} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad \text{s.t. } \mathbf{A}, \mathbf{S} \geq 0. \quad (3)$$

The objective function of NMF is convex with respect to either \mathbf{A} or \mathbf{S} but not both, making the solution space very huge. Therefore, various *priors* are embedded into factor matrices to reduce the searching space, so as to achieve more interpretable factorization. Among them, abundance sparsity is commonly used due to the fact not all the endmembers but a subset of endmembers contribute to a pixel. Imposing the sparsity constraint on \mathbf{S} , the sparse NMF, in general, takes the following form:

$$\min_{\mathbf{A}, \mathbf{S}} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \mathcal{R}(\mathbf{S}) \quad \text{s.t. } \mathbf{A}, \mathbf{S} \geq 0 \quad (4)$$

where $\mathcal{R}(\mathbf{S})$ is a sparsity-inducing regularization and $\lambda \geq 0$ balances the data fidelity and sparse regularization.

The sparse regularizations can be broadly divided into convex and nonconvex ones. L_1 norm earns its popularity among the convex regularizations as it yields sparse solutions very effectively. However, L_1 sparsity violates abundance sum-to-one constraint and always causes underestimation of target abundances, dramatically limiting its practical utility. Beyond convex regularizations, nonconvex sparse ones such as $L_{1/2}$ norm [26], reweighted L_1 norm [23], and log-sum penalty [17] have shown better sparsity-inducing ability by producing less-biased estimations. Nevertheless, nonconvex penalties promote sparsity regardless of the convexity of NMF with respect to individual factor matrix, making the optimization more intractable. This possibly conflicts with their initial goal of shrinking the solution space of NMF. Therefore, it is necessary to find a new sparsity-promoting regularization which simultaneously maintains the convexity of NMF with respect to individual factor matrix and avoids underestimation of target abundances.

III. GMC-NMF UNMIXING MODEL

In this section, we introduce the proposed GMC-regularized NMF for hyperspectral unmixing and address the associated implementation and optimization issues.

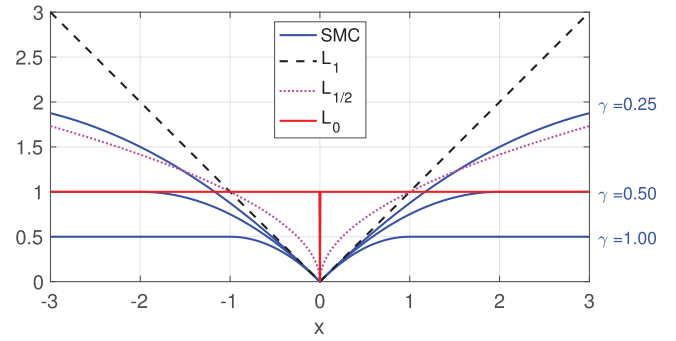


Fig. 1. Scaled minimax-concave penalty with respect to γ values.

A. GMC-NMF

Before a comprehensive introduction of the GMC penalty, we first analyze its unbiased sparsity-inducing ability based on its scalar version, i.e., scaled minimax-concave (SMC) penalty.

Definition 1: Parameterized by a positive scalar γ , the scaled minimax-concave penalty function $\phi_\gamma : \mathbb{R} \rightarrow \mathbb{R}$ is defined as

$$\phi_\gamma(x) = \begin{cases} |x| - \frac{\gamma}{2}x^2, & |x| \leq \frac{1}{\gamma} \\ \frac{1}{2\gamma}, & |x| \geq \frac{1}{\gamma} \end{cases} \quad (5)$$

SMC penalty can be identically represented as a Moreau envelope [36], i.e.,

$$\phi_\gamma(x) = |x| - \min_{v \in \mathbb{R}} \left\{ |v| + \frac{\gamma}{2}(x - v)^2 \right\}. \quad (6)$$

Fig. 1 plots the SMC penalty for different values of γ . It can be observed from the figure that first, the SMC penalty is nonconvex and γ controls its nonconvex level, i.e., larger γ indicates higher nonconvexity. Specially, when γ approaches zero, the SMC penalty gradually degrades to L_1 sparse penalty. Second, SMC is able to promote sparsity. Like L_1 and $L_{1/2}$ penalty, the SMC penalty is not differentiable when $x = 0$. As discussed in [37], zero becomes the only solution when the regularization parameter is large enough. Third, SMC penalty produces unbiased sparse solution. The gradient of GMC penalty decreases with the increasing of x , meaning that large-amplitude components are less likely to be shrunk to 0. This is different from L_1 penalty which treats all the values equally as its gradient is constant for $x \neq 0$.

Based on SMC, the GMC penalty is an extended version for multivariate cases whose definition is as follows.

Definition 2: Given $\mathbf{B} \in \mathbb{R}^{M \times N}$ and $\mathbf{x} \in \mathbb{R}^{N \times 1}$ and $\lambda \geq 0$. GMC penalty function: $\mathbb{R}^{N \times 1} \rightarrow \mathbb{R}$ is defined as

$$\mathcal{R}_{\mathbf{B}}(\mathbf{x}) = \|\mathbf{x}\|_1 - \min_{\mathbf{v} \in \mathbb{R}^{N \times 1}} \left\{ \|\mathbf{v}\|_1 + \frac{1}{2} \|\mathbf{B}(\mathbf{x} - \mathbf{v})\|_F^2 \right\}. \quad (7)$$

Different from most sparse penalties which are additively separable for each component of a vector, GMC penalty is nonseparable, i.e., $\mathcal{R}_{\mathbf{B}}(\mathbf{x}) \neq \sum_n \phi_{\gamma_n}(x_n)$ when $\mathbf{B}^T \mathbf{B}$ is non-diagonal. This property also makes the GMC penalty capable of maintaining the convexity of regularized objective function when \mathbf{B} is appropriately selected.

Lemma 1: Let $\mathbf{y} \in \mathbb{R}^{M \times 1}$, $\mathbf{A} \in \mathbb{R}^{M \times N}$ and $\lambda \geq 0$. $F: \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}$

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_F^2 + \lambda \mathcal{R}_{\mathbf{B}}(\mathbf{x}) \quad (8)$$

is convex provided that $\mathbf{B} = \sqrt{\frac{\gamma}{\lambda}} \mathbf{A}$, where $0 \leq \gamma \leq 1$.

Proof:

$$\begin{aligned} F(\mathbf{x}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_F^2 + \lambda \|\mathbf{x}\|_1 - \min_{\mathbf{v}} \left\{ \lambda \|\mathbf{v}\|_1 + \frac{\gamma}{2} \|\mathbf{A}(\mathbf{x} - \mathbf{v})\|_F^2 \right\} \\ &= \max_{\mathbf{v}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_F^2 + \lambda \|\mathbf{x}\|_1 - \lambda \|\mathbf{v}\|_1 - \frac{\gamma}{2} \|\mathbf{A}(\mathbf{x} - \mathbf{v})\|_F^2 \right\} \\ &= \max_{\mathbf{v}} \left\{ \frac{\mathbf{x}^T}{2} (\mathbf{A}^T \mathbf{A} - \gamma \mathbf{A}^T \mathbf{A}) \mathbf{x} + \lambda \|\mathbf{x}\|_1 + g(\mathbf{x}, \mathbf{v}) \right\} \\ &= \frac{\mathbf{x}^T}{2} (\mathbf{A}^T \mathbf{A} - \gamma \mathbf{A}^T \mathbf{A}) \mathbf{x} + \lambda \|\mathbf{x}\|_1 + \max_{\mathbf{v}} \{g(\mathbf{x}, \mathbf{v})\} \end{aligned} \quad (9)$$

where $g(\mathbf{x}, \mathbf{v})$ is affine in \mathbf{x} and convex since it is the point-wise maximum of a set of convex functions. As $0 \leq \gamma < 1$, $\mathbf{A}^T \mathbf{A} - \gamma \mathbf{A}^T \mathbf{A}$ is positive semidefinite, meaning that $\frac{\mathbf{x}^T}{2} (\mathbf{A}^T \mathbf{A} - \gamma \mathbf{A}^T \mathbf{A}) \mathbf{x}$ is convex. Considering convex nature of $\|\mathbf{x}\|_1$, $F(\mathbf{x})$ is convex. ■

In summary, GMC penalty is able to produce unbiased sparsity while simultaneously maintaining the convexity of the regularized problem. This makes the GMC-regularized sparse problem share the favorable merits of convex optimization such as unique minimum and robustness while producing stronger sparsity than the standard convex sparse regularization options. Thanks to these advantages, GMC penalty has been successfully applied to signal denoising [38] and low-rank approximations [35].

As discussed earlier, sparsity is an important *prior* for abundance and is beneficial to limit the solution space of NMF. Considering the attractive properties of the GMC penalty in strong sparsity promoting and convexity preserving, we introduce this penalty to abundance matrix for robust unmixing. Replacing $\mathcal{R}(\mathbf{S})$ in (4) with the GMC penalty, we have a new sparse NMF unmixing model abbreviated as GMC-NMF whose objective function can be derived as follows:

$$F(\mathbf{A}, \mathbf{S}, \mathbf{V}) = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_1 - \min_{\mathbf{V}} \left\{ \lambda \|\mathbf{V}\|_1 + \frac{\gamma}{2} \|\mathbf{A}(\mathbf{S} - \mathbf{V})\|_F^2 \right\} \quad \text{s.t. } \mathbf{A}, \mathbf{S} \geq 0. \quad (10)$$

Unlike separable L_1 and $L_{1/2}$ penalties, \mathbf{A} is coupled with both data fidelity and sparse-inducing terms. More importantly, \mathbf{A} captures the relationship among the elements in abundance maps. As \mathbf{A} represents the endmember matrix, the sparse unmixing model in (10) actually simultaneously builds the semantical correlation between the abundance distribution of different materials. This is consistent with the fact that the abundance maps of different materials such as trees and grass should be correlated, thereby facilitating more accurate abundances estimation, which in turn benefits endmember extraction.

B. Optimization Procedure

The minimization problem of $F(\mathbf{A}, \mathbf{S}, \mathbf{V})$ can be rewritten as a saddle-point problem, i.e.,

$$\arg \min_{\mathbf{A}, \mathbf{S}} \max_{\mathbf{V}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_1 - \lambda \|\mathbf{V}\|_1 - \frac{\gamma}{2} \|\mathbf{A}(\mathbf{S} - \mathbf{V})\|_F^2. \quad (11)$$

The update rule of each variable can be derived separately. Fixing \mathbf{S} and \mathbf{V} , the subproblem of \mathbf{A} is

$$\arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 - \frac{\gamma}{2} \|\mathbf{A}(\mathbf{S} - \mathbf{V})\|_F^2. \quad (12)$$

The updated rule of \mathbf{A} is deduced under the framework of Lagrange multipliers. The Lagrange function Φ of \mathbf{A} is as follows:

$$\Phi = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 - \frac{\gamma}{2} \|\mathbf{A}(\mathbf{S} - \mathbf{V})\|_F^2 + \text{Tr}(\Gamma \mathbf{A}) \quad (13)$$

where Γ is the Lagrange multiplier of \mathbf{A} . Taking the partial derivatives of Φ with respect to \mathbf{A} , we get

$$\mathbf{X}\mathbf{S}^T - \mathbf{A}\mathbf{S}\mathbf{S}^T - \gamma \mathbf{A}(\mathbf{S} - \mathbf{V})(\mathbf{S} - \mathbf{V})^T + \Gamma = \mathbf{0}. \quad (14)$$

Defining $\mathbf{D} = \mathbf{S}\mathbf{S}^T + \gamma(\mathbf{S} - \mathbf{V})(\mathbf{S} - \mathbf{V})^T$, (14) can be identically expressed as

$$\mathbf{X}\mathbf{S}^T - \mathbf{A}\mathbf{D} = \mathbf{0}. \quad (15)$$

According to Karush–Kuhn–Tucker conditions, the following equation is met:

$$\Gamma .* \mathbf{A} = \mathbf{0} \quad (16)$$

in which $.*$ denotes element-wise multiplication.

Similar to [39], by introducing $\mathbf{D} = \mathbf{D}^+ - \mathbf{D}^-$ where $\mathbf{D}^+ = (|\mathbf{D}| + \mathbf{D})/2$ and $\mathbf{D}^- = (|\mathbf{D}| - \mathbf{D})/2$ and substituting (16) into (15), the update rule of \mathbf{A} can be derived as

$$\mathbf{A} \leftarrow \mathbf{A} .* (\mathbf{X}\mathbf{S}^T + \mathbf{A}\mathbf{D}^-) ./ (\mathbf{A}\mathbf{D}^+). \quad (17)$$

As for the minimization problems of \mathbf{S} and \mathbf{V} , both of them are convex and can be decomposed into two parts, i.e., one of which is differentiable and the other is nondifferentiable with a fast proximal operator. Forward–backward splitting algorithm is an effective tool to solve such problems. For the forward step, taking \mathbf{S} for example, gradient decent with step size α can be used to get

$$\mathbf{S}' \leftarrow \mathbf{S} - \alpha (\mathbf{A}^T (\mathbf{A}\mathbf{S} - \mathbf{X}) - \gamma \mathbf{A}^T \mathbf{A} (\mathbf{S} - \mathbf{V})). \quad (18)$$

Afterwards, backward step applies the proximal operator on \mathbf{S}' to yield the update rule for \mathbf{S} :

$$\mathbf{S} \leftarrow \max(\mathbf{S}' - \lambda, 0) \quad (19)$$

where \max is a proximal function guaranteeing the nonnegativity of \mathbf{S} .

Similarly, \mathbf{V} is first updated using gradient decent in the forward step, resulting in

$$\mathbf{V}' \leftarrow \mathbf{V} - \alpha \gamma \mathbf{A}^T \mathbf{A} (\mathbf{S} - \mathbf{V}). \quad (20)$$

Algorithm 1: GMC-NMF-Based Unmixing Procedure.**Input:** Hyperspectral image \mathbf{X} , parameters λ and γ ;**Output:** \mathbf{A} , \mathbf{S} , and \mathbf{V} .

```

1: Initialize  $\mathbf{A}$ ,  $\mathbf{S}$ ,  $\mathbf{V}$ 
2: repeat
3:   Update  $\mathbf{A}$  with (15)
4:   repeat
5:     Forward step:
6:       Obtain  $\mathbf{S}'$  with (18)
7:       Obtain  $\mathbf{V}'$  with (20)
8:     Backward step:
9:       Update  $\mathbf{S}$  with (19)
10:      Update  $\mathbf{V}$  with (21)
11:   until The stopping condition is met.
12: until GMC-NMF converges

```

Subsequently, the soft shrinkage function is applied on \mathbf{V}' to obtain the update rule of \mathbf{V} , i.e.,

$$\mathbf{V} \leftarrow \begin{cases} v - \lambda, & v \geq \Delta \\ v + \lambda, & v \leq -\Delta \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where $v \in \mathbb{R}$ is an element in \mathbf{V} . Selesnick [32] has suggested setting $0 < \alpha < 2/\rho$, where $\rho = \max\{1, \gamma/(1 - \gamma)\} \|\mathbf{A}^T \mathbf{A}\|_2$. Due to the convex nature of solving \mathbf{S} and \mathbf{V} , α theoretically has no impact on the solution accuracy. Generally, larger α contributes to accelerating the convergence; therefore, α is set to $1.9/\rho$ in practice.

C. Algorithm Implementation

Algorithm 1 summarizes the optimization procedure of the proposed unmixing method. Here, we address the following significant issues regarding the implementation of the algorithm.

Algorithm initialization: The initialization of \mathbf{A} , \mathbf{S} , and \mathbf{V} is important due to the nonconvex nature of (10). Generally, non-random initialization can produce more favorable unmixing results than random initialization. To this end, we use VCA-FCLS (fully constrained least squares) to initialize the optimization procedure. Specifically, we first use VCA [7] to obtain the initial estimation of \mathbf{A} . Afterwards, \mathbf{S} is assigned with the abundances obtained by FCLS [8]. As for \mathbf{V} , it is experimentally set as $\mathbf{V} = \mathbf{S}$.

ASC Constraint: ASC constraint is helpful to reduce the searching space of optimization. As in [26], we satisfy ASC constraint by augmenting the observed data matrix \mathbf{X} and end-member matrix \mathbf{A} as

$$\begin{aligned} \mathbf{X} &\leftarrow \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}_N^T \end{bmatrix} \\ \mathbf{A} &\leftarrow \begin{bmatrix} \mathbf{A} \\ \delta \mathbf{1}_R^T \end{bmatrix} \end{aligned} \quad (22)$$

where δ controls the influence of the ASC constraint. As suggested in [16], we set $\delta = 15$.

Endmember Number: The number of endmembers R indicates the dimensionality spanned by endmembers. On the

synthetic data, the number is the same as the one used to produce the data. As for the real-world data, we adopt hyperspectral signal subspace identification by minimum error (Hysime) [40] to estimate R .

Convergence Condition: Since (10) does not have a simple explicit formula, it is difficult to calculate the detailed numerical value. Instead, we define two stopping conditions to check whether the optimization converges. By defining $L^{(k)} = \frac{1}{2} \|\mathbf{X} - \mathbf{A}^{(k)} \mathbf{S}^{(k)}\|_F^2$, where k denotes the iteration number, the first condition is that the relative reconstruction error of \mathbf{X} is less than 10^{-4} , i.e.,

$$\frac{\text{abs}(L^{(k+1)} - L^{(k)})}{L^{(k)}} < 10^{-4}. \quad (23)$$

The second one is that the maximum iteration number exceeds a preset number of 3000. The optimization terminates when either condition is met.

IV. EXPERIMENTS

In this section, we compare proposed GMC-NMF against a series of alternative unmixing approaches on both synthetic data and real-world data to thoroughly demonstrate its effectiveness. Specifically, $L_{1/2}$ -NMF [26], NMF, TV-RSNMF [23], Arctan-NMF [41], and VCA-FCLS are selected. Their parameters are automatically set as suggested values in the original implementation.

A. Performance Evaluation Criteria

The spectral angle distance (SAD) and root-mean-square error (RMSE) were used to qualify the unmixing performance. SAD computes the similarity between reference endmembers \mathbf{A}_r and corresponding estimation $\hat{\mathbf{A}}_r$, whose definition is

$$\text{SAD}_r = \arccos \left(\frac{\mathbf{A}_r^T \hat{\mathbf{A}}_r}{\|\mathbf{A}_r\| \|\hat{\mathbf{A}}_r\|} \right). \quad (24)$$

RMSE describes the error between ground-truth abundance map and estimated map, which is defined as

$$\text{RMSE}_r = \left(\frac{1}{N} |\mathbf{S}_r - \hat{\mathbf{S}}_r|^2 \right)^{\frac{1}{2}} \quad (25)$$

where $\hat{\mathbf{S}}_r$ is the estimated abundance map of the r th endmember.

B. Experiments on Synthetic Data

The synthetic data is produced following the method provided in [42]. Six pure signatures (Carnallite, Ammonio-jarosite, Almandine, Brucite, Axinite, and Chlonte) containing 224 bands with wavelengths in the range of 0.38–2.5 μm are first selected from the USGS library [43] to generate endmembers. Fig. 2 shows the spectral signatures of these endmembers. Afterwards, the abundances are yielded as follows: 1) A synthetic image size of $z^2 \times z^2$ is partitioned into z^2 disconnected regions, each of which covers $z \times z$ pixels. 2) Two randomly selected endmembers are used to fill each pixel of a region, with ratios set to β and $1 - \beta$, respectively. 3) Each abundance map is convolved with a Gaussian filter whose variance is set to 2, producing highly mixed data. 4) The fractions of all endmembers

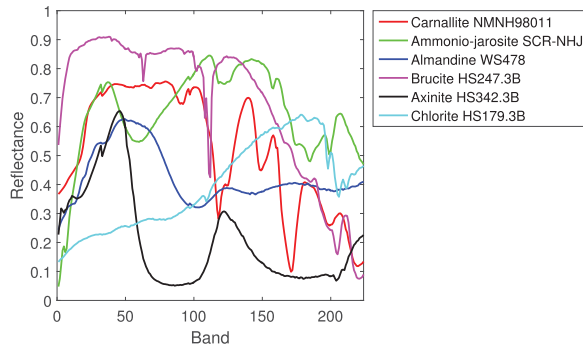


Fig. 2. Spectral signatures of six endmembers used in the synthetic data.

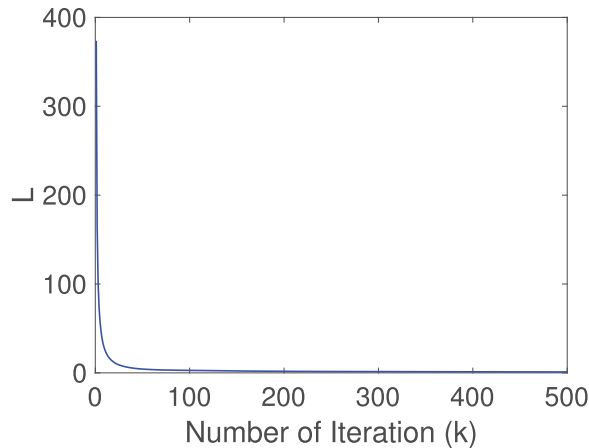


Fig. 3. Convergence curve of GMC-NMF.

in each pixel are rescaled to meet the ASC constraint. Finally, the synthetic data is generated using the above abundances and endmembers.

In order to test the sensitivity of methods to noise, the generated clean HSI is also degraded by zero-mean additive Gaussian noise whose SNR is defined as

$$\text{SNR} = 10 \log_{10} \frac{E[\mathbf{y}^T \mathbf{y}]}{E[\mathbf{e}^T \mathbf{e}]} \quad (26)$$

where \mathbf{y} and \mathbf{e} are the clean signal and the noise at a pixel and $E[\cdot]$ denotes the expectation operator.

1) *Convergence Analysis*: This experiment aims to characterize the convergence property of our method. The synthetic data was generated under the configuration of $z = 8$, $\beta = 0.8$, $\text{SNR} = \infty$. As mentioned earlier, it is rather difficult to get the explicit objective function of GMC-NMF. Alternatively, we plot the values of reconstruction error, i.e., $L^{(k)} = \frac{1}{2} \|\mathbf{X} - \mathbf{A}^{(k)} \mathbf{S}^{(k)}\|_F^2$ to present its convergence property. As shown in Fig. 3, the values drop quickly before iteration 100 and gradually decrease thereafter until convergence.

2) *Parameter Setting*: The synthetic data was produced by setting $z = 8$, $\beta = 0.8$, $\text{SNR} = 25$ dB in this experiment. λ and γ are two parameters controlling the sparsity-inducing capability of GMC regularizer. In the experiments, we set λ in the set of $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 3, 5\}$ and change γ in the range of $\{0.02, 0.04, 0.06, 0.08, 0.1, 0.3, 0.5, 0.7, 0.9\}$.

Fig. 4 shows the unmixing performance with respect to all the combinations of two parameters. Except very small values of γ , both SAD and RMSE almost stay steady regardless of the settings of λ . The main reason is that GMC regularizer is able to preserve the convexity of original objective function, making the optimization less possible trap in local minimal and less sensitive to parameter settings. When two parameters are close to 0, the RMSE increases, implying the positive effect of GMC sparsity regularizer. Moreover, as the growth of γ is in the range of 0.1–0.9, the corresponding SAD and RMSE first slightly increase, and then sharply when $\gamma \geq 0.7$. Since the promising SAD and RMSE are achieved at $\gamma = 0.1$, this value is suggested for more favorable unmixing accuracy. Except very small γ , both SAD and RMSE are robust to λ settings. Based on the above observation, we set $\lambda = 1$ in the following experiments.

3) *Robustness to Noise*: Hyperspectral images are easily corrupted in the acquisition. This experiment aims to evaluate the influence of different levels of noises on unmixing. The clean HSI was generated under the setting of $z = 8$ and $\beta = 0.8$. We then added different levels of SNR, i.e., 15, 20, 25, and 30 dB, and ∞ (noise free), to the clean HSI to produce noisy HSIs.

Fig. 5 shows a comparative unmixing performances of all the methods. Overall, most unmixing methods show descending trend as the SNR increases. The limited consideration of sparsity property of abundances maps results in the worst SAD and RMSE provided by VCA-FCLS. Theoretically, sparse representation has the capacity of filtering the noise, making the methods less affected by the noises. Thanks to the sparsity-promoting ability of $L_{1/2}$ and arctan regularization, better results are achieved especially in the noisy cases. Moreover, with the ability of inducing unbiased sparsity while simultaneously preserving convexity, GMC-NMF offers the most promising unmixing results in both SAD and RMSE. This experiment further verifies the effectiveness of GMC-NMF in unmixing.

4) *Generalization to Mixing Levels*: β is an indicator of mixing degree, i.e., larger values of β imply less mixing degree. In the experiment, the pixel size and SNR were, respectively, fixed as 64×64 and 25 dB. Varying β from 0.5 to 1 with an interval of 0.1, Fig. 6 shows the unmixing ability of all the competing methods. Overall $L_{1/2}$ -NMF and GMC-NMF dominate among all the methods because of more capability of promoting sparser abundances. Compared with $L_{1/2}$ -NMF, GMC-NMF performs better, especially in RMSE. The main reason is that GMC regularization induces sparsity and preserves the attractive convex characteristic of NMF with respect to each variable at the same time, better helping NMF get rid of many local minimal.

5) *Running Time Comparison*: All the methods were run on a Windows 10 machine equipped with Intel(R) Xeon(R) Silver 4114 CPU@2.20 GHz and 32 GB RAM. The running time is shown in Table I. Except TV-RSNMF, GMC-NMF requires the most running time due to additional computational cost for unbiased sparse solutions. In most scenarios, unmixing is performed offline and it is worthwhile to spend more time improving the unmixing accuracy. Moreover, thanks to the graphics processing unit and parallel computing, the running time can be significantly shortened.

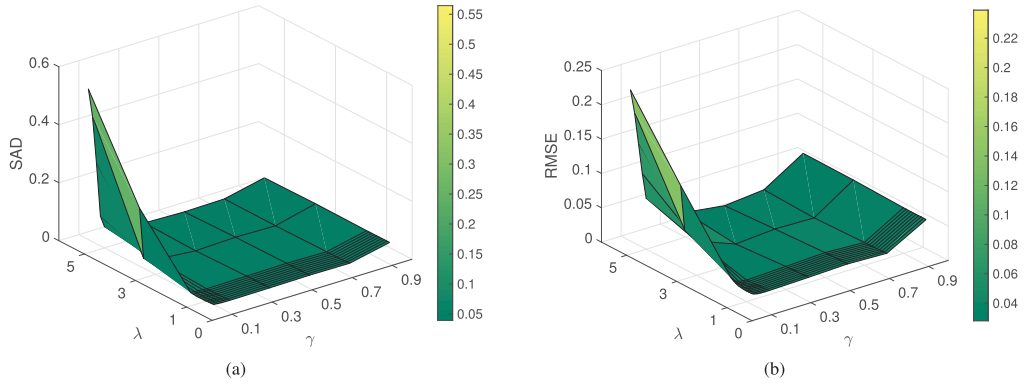


Fig. 4. Performance of GMC-NMF with respect to λ and γ settings (best viewed by zooming in).

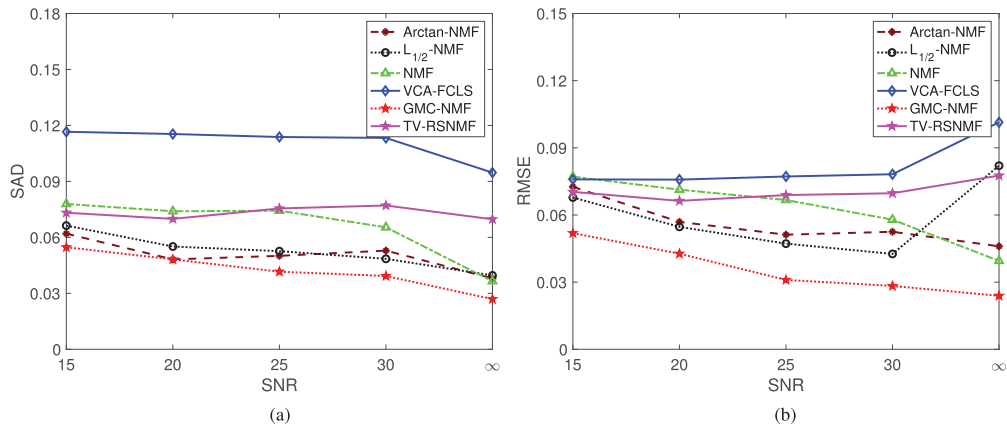


Fig. 5. SAD and RMSE with respect to different noise levels.

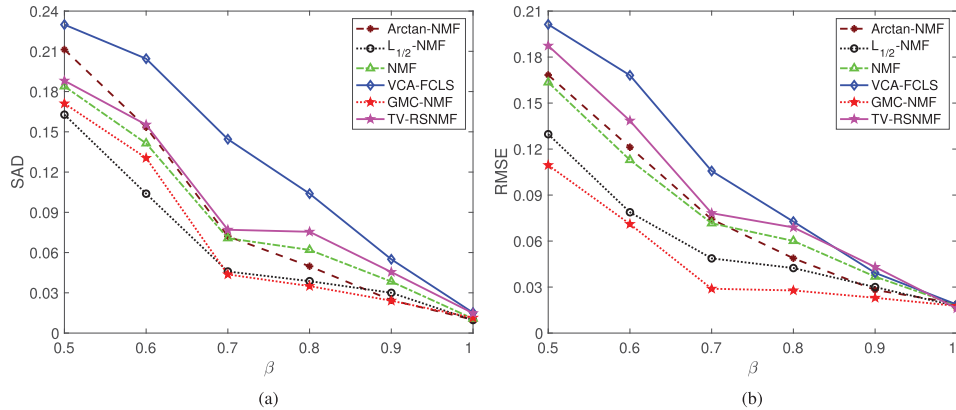


Fig. 6. SAD and RMSE with respect to different mixing levels.

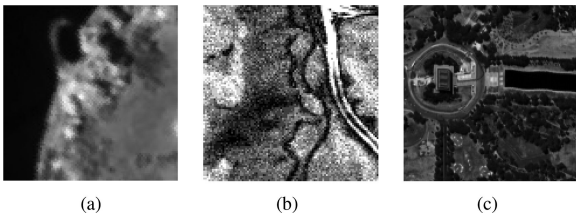


Fig. 7. Three real-world datasets shown in gray. (a) 140th band image of Samson. (b) Third band image of Jasper Ridge. (c) 150th band image of Washington DC Mall.

TABLE I
RUNNING TIME OF ALL THE METHODS

Algorithm	VCA-FCLS	NMF	$L_{1/2}$ -NMF	Arctan-NMF	TV-RSNMF	GMC-NMF
Time (s)	2.15	1.51	0.87	1.17	32.22	25.43

C. Experiments on Real-World Data

In this section, we evaluate all the methods on three real-world hyperspectral images for more comprehensive comparison.

TABLE II
MEANS AND STANDARD DEVIATIONS OF THE SAD ON SAMSON DATA

Algorithm	VCA-FCLS	NMF	$L_{1/2}$ -NMF	Arctan-NMF	TV-RSNMF	GMC-NMF
Soil	0.0252±0.33%	0.0236±0.53%	0.0236±0.52%	0.0185±0.20%	0.0184±0.70%	0.0328±0.86%
Tree	0.0408±0.6%	0.0540±0.44%	0.0518±0.38%	0.0428±0.09%	0.0483±0.28%	0.0301±0.04%
Water	0.3240±14.97%	0.0978±1.48%	0.1079±1.57%	0.2144±8.39%	0.1726±0.79%	0.0892±2.37%
Mean	0.1300±5.06%	0.0585±0.44%	0.0611±0.58%	0.0919±2.80%	0.0798±0.33%	0.0507±1.58%

The entries in bold imply best unmixing performance.

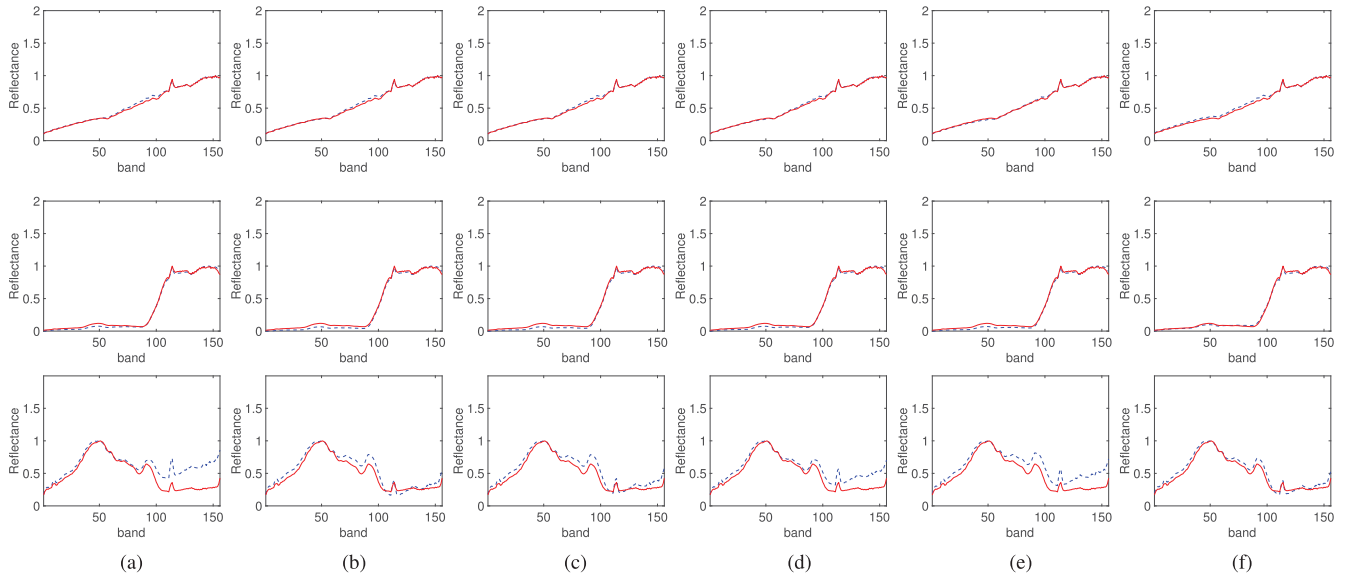


Fig. 8. Estimated endmembers by six unmixing methods on the Samson dataset. From top to bottom, the rows are the spectral signatures of soil, tree, and water, respectively. Solid lines denote the reference endmembers and dashed lines denote the estimated endmembers. (a) VCA-FCLS. (b) NMF. (c) $L_{1/2}$ -NMF. (d) Arctan-NMF. (e) TV-RSNM. (f) GMC-NMF.

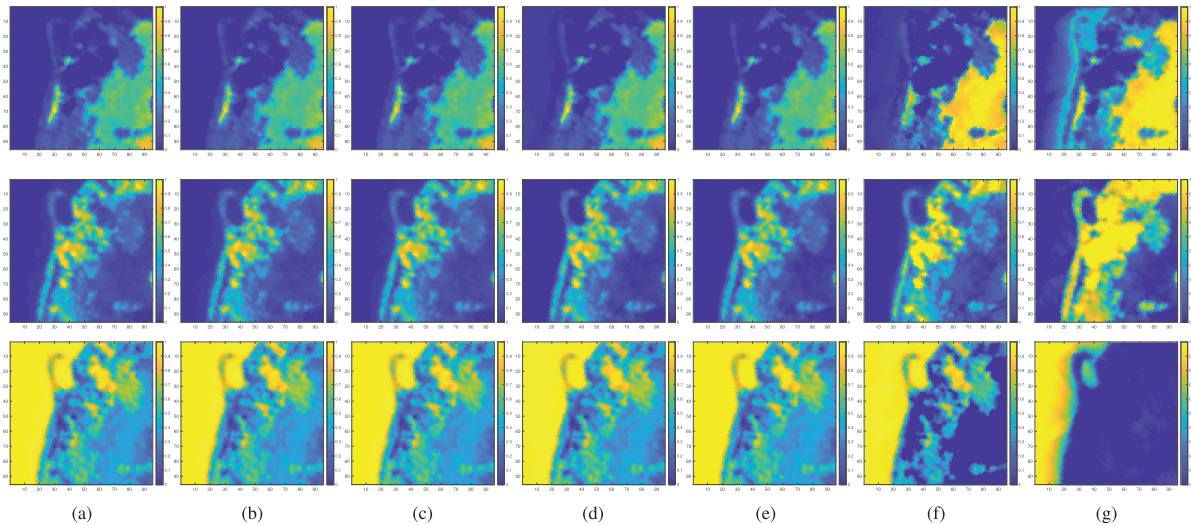


Fig. 9. Estimated abundance maps by six unmixing methods on the Samson dataset. From top to bottom, the rows are the abundance maps of soil, tree, and water, respectively. (a) VCA-FCLS. (b) NMF. (c) $L_{1/2}$ -NMF. (d) Arctan-NMF. (e) TV-RSNMF. (f) GMC-NMF. (g) Reference.

1) *Samson Dataset*: The Samson dataset is an HSI containing 952×952 pixels and 156 spectral bands ranging from 0.4 to $0.89 \mu\text{m}$.¹ A subscene size of 95×95 is cropped

¹[Online]. Available: Link:<http://opticks.org/confluence/display/opticks/Sample+Data>

for the experiments, whose 140th band is shown in Fig. 7(a). Previous studies [41], [44] suggest that there should be three kinds of targets of interest, i.e., soil, tree, and water, respectively. Table II presents a quantitative comparison of all the methods on SAD criteria. The proposed GMC-NMF offers the best mean SAD. Fig. 8 shows a qualitative comparison with endmember

TABLE III
MEANS AND STANDARD DEVIATIONS OF THE SAD ON JASPER RIDGE DATA

Algorithm	VCA-FCLS	NMF	$L_{1/2}$ -NMF	Arctan-NMF	TV-RSNMF	GMC-NMF
Tree	0.2454±8.65%	0.1591±8.72%	0.0585±4.05%	0.1464±7.15%	0.0621±4.10%	0.1048±1.88%
Water	1.0442±3.51%	0.2549±2.77%	0.1900±5.43%	0.3566±11.89%	0.2258±6.25%	0.0934±3.42%
Soil	0.0667±2.38%	0.3258±28.68%	0.1007±12.08%	0.1509±9.46%	0.1069±13.67%	0.1854±17.36%
Road	0.1997±4.06%	0.1902±19.59%	0.2255±31.75%	0.2441±35.76%	0.2686±34.95%	0.1054±7.30%
Mean	0.3890±2.67%	0.2325±8.86%	0.1437±9.37%	0.2245±11.28%	0.1658±11.17%	0.1230±5.61%

The entries in bold imply best unmixing performance.

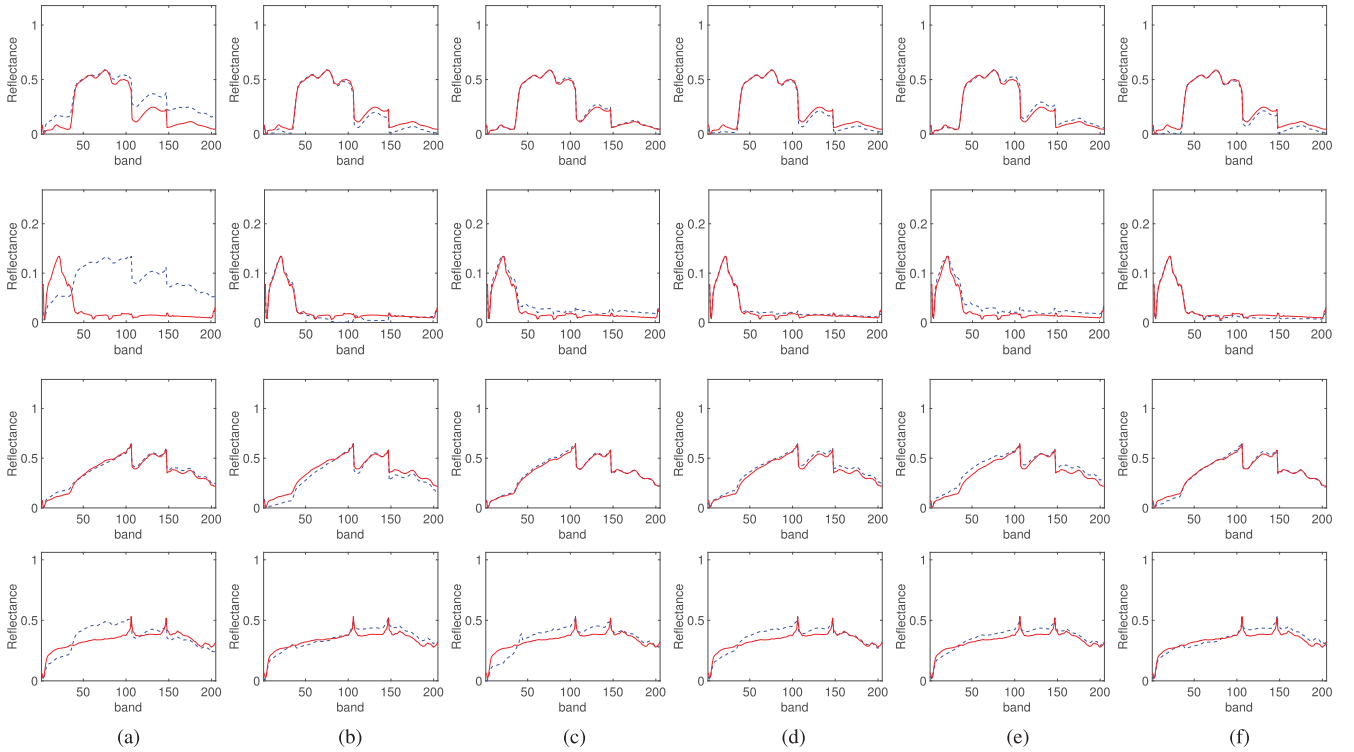


Fig. 10. Estimated endmembers by six unmixing methods on the Jasper Ridge dataset. From top to bottom, the rows are the spectral signatures of tree, water, soil, and road, respectively. Solid lines denote the reference endmembers and dashed lines denote the estimated endmembers. (a) VCA-FCLS. (b) NMF. (c) $L_1=2$ -NMF. (d) Arctan-NMF. (e) TV-RSNMF. (f) GMC-NMF.

TABLE IV
MEANS AND STANDARD DEVIATIONS OF THE SAD ON WASHINGTON DC MALL DATA

Algorithm	VCA-FCLS	NMF	$L_{1/2}$ -NMF	Arctan-NMF	TV-RSNMF	GMC-NMF
Tree	0.1995±0.87%	0.2153±0.39%	0.2116±0.06%	0.2045±0.46%	0.20711±1%	0.2151±0.62%
Grass	0.3047±8.61%	0.2939±7.01%	0.2918±6.20%	0.2285±1.66%	0.2961±4.34%	0.2826±5.55%
Street	0.3875±3.69%	0.3433±3.72%	0.3452±2.73%	0.4576±1.46%	0.3338±1.57%	0.3283±1.78%
Roof	0.2002±0.05%	0.2143±2.60%	0.1991±4.38%	0.1668±0.49%	0.2163±2.84%	0.2079±3.55%
Water	0.0444±0.04%	0.0707±1.07%	0.0589±0.02%	0.0431±0.28%	4.66±0.53%	0.0530±0.71%
Mean	0.2273±0.04%	0.2275±1.92%	0.2212±0.93%	0.2201±0.56%	0.2200±0.51%	0.2174±0.74%

The entries in bold imply best unmixing performance.

signatures generated by all the competing methods and those of reference endmembers. As can be seen, the endmembers extracted by the proposed method are more accurate by closely matching the reference ones, especially for water target. The experimental results provide clear support for the effectiveness of GMC-NMF in unmixing. Fig. 9 visualizes the abundance maps generated by all the methods. The abundance maps produced by GMC-NMF are closer to the reference ones and

go beyond alternative methods, confirming that our method is effective in unmixing.

2) *Jasper Ridge Dataset*: Jasper Ridge dataset is a widely used HSI in the unmixing task. It was acquired by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor, consisting of 512×614 pixels and 224 bands from 380 to 2500 nm. After removing the bands corrupted by water vapor, i.e., bands 1108–1112 and 154–166, we used a subimage with

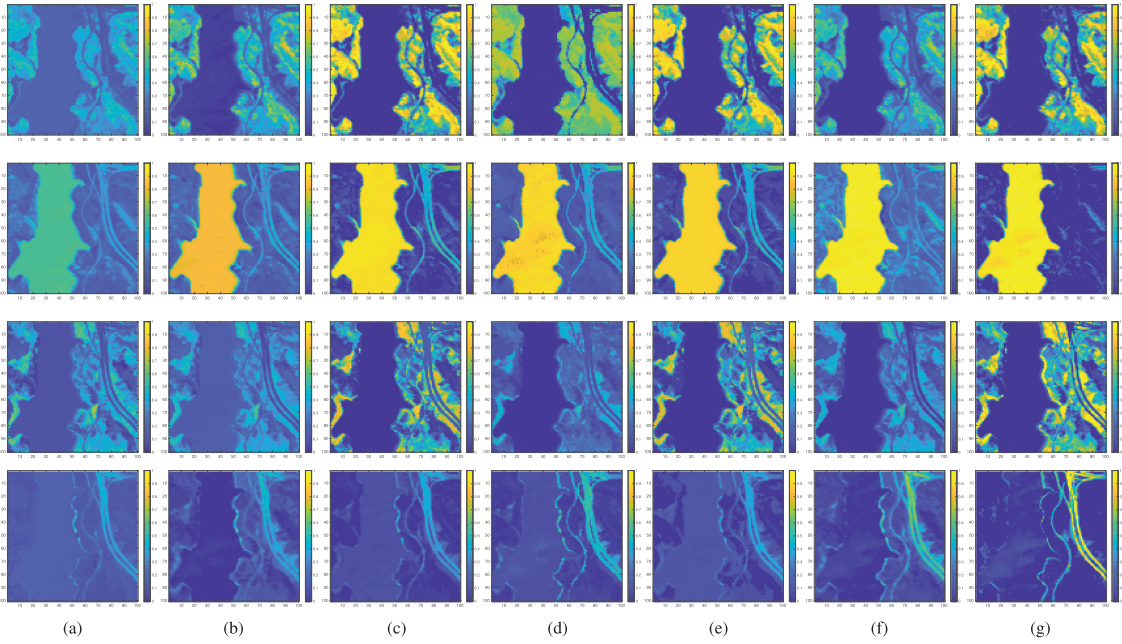


Fig. 11. Estimated abundance maps by six unmixing methods on the Jasper Ridge dataset. From top to bottom, the rows are the abundance maps of tree, water, soil, and road, respectively. a) VCA-FCLS. (b) NMF. (c) $L_{1/2}$ -NMF. (d) Arctan-NMF. (e) TV-RSNMF. (f) GMC-NMF. (g) Reference.

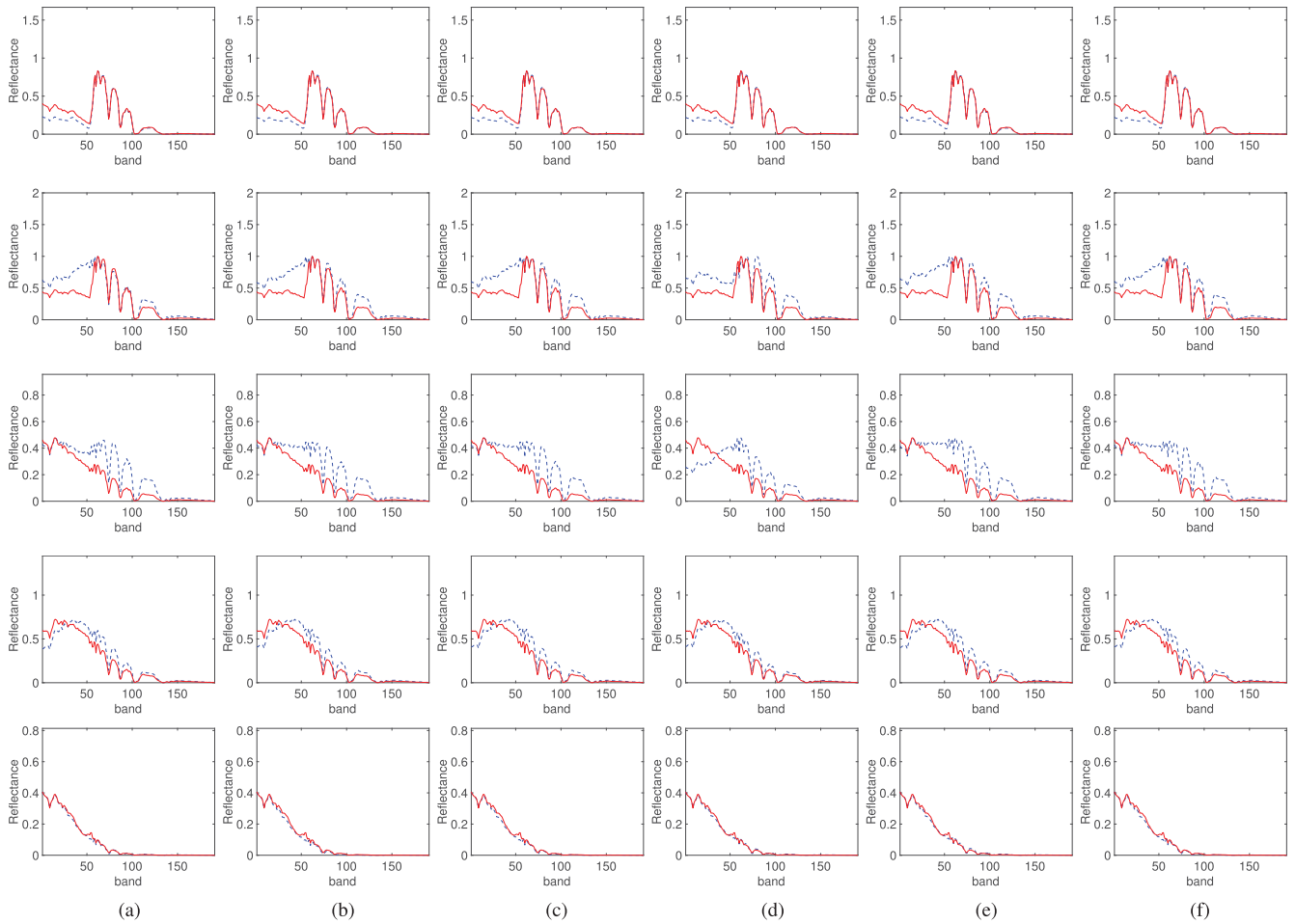


Fig. 12. Estimated endmembers by six unmixing methods on the Washington DC Mall dataset. From top to bottom, the rows are the spectral signatures of tree, grass, street, roof, and water, respectively. Solid lines denote the reference endmembers and dashed lines denote the estimated endmembers. (a) VCA-FCLS. (b) NMF. (c) $L_{1/2}$ -NMF. (d) Arctan-NMF. (e) TV-RSNMF. (f) GMC-NMF.

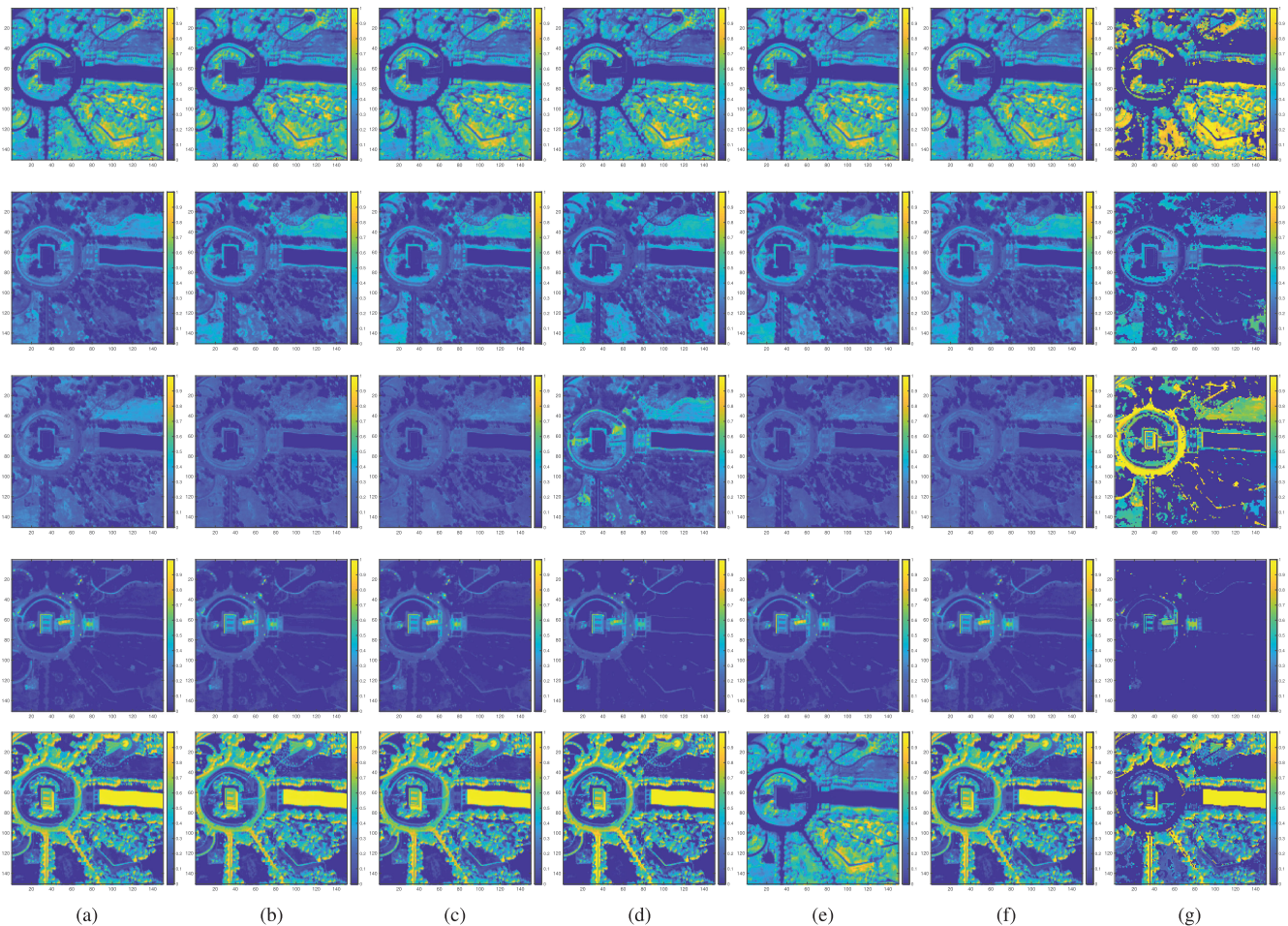


Fig. 13. Estimated abundance maps by six unmixing methods on the Washington DC Mall dataset. From top to bottom, the rows are the abundance maps of tree, grass, street, roof, and water, respectively. (a) VCA-FCLS. (b) NMF. (c) $L_{1/2}$ -NMF. (d) Arctan-NMF. (e) TV-RSNMF. (f) GMC-NMF. (g) Reference.

205 bands and 100×100 pixels for experimental evaluation. Fig. 7(b) shows the third band image from the figure, the dataset contains heavy noises, which threat challenges to unmixing. In the experiment, we set four kinds of materials as endmembers, including tree, water, soil, and road.

Table III presents the quantitative comparison of all the competing methods with respect to SAD. As the data contains heavy noises, VCA-FCLS provides the worst SADs. The sparse regularization is beneficial to reduce the solution space of NMF and suppress the negative effect of noises, yielding better SADs of $L_{1/2}$ -NMF, Arctan-NMF, TV-RSNMF, and GMC-NMF. Thanks to the inherent merits of promoting unbiased sparsity with maintained convexity of original problem, even better results are achieved when using GMC-NMF. The endmembers produced by all the methods are shown in Fig. 10. Superior results are seen for GMC-NMF by providing the best-matched signatures. Fig. 11 illustrates a visual comparison resulted in abundances. It is worth mentioning that our method also gives very competitive results and provides the most effective results on water and road.

3) *Washington DC Mall Dataset*: This dataset was collected by the urban hyperspectral digital imagery collection experiment sensor over the Washington DC Mall. The original raw

image contains 1208×307 pixels and 210 bands, but we used a cropped subimage of 150×150 pixels. Moreover, the low SNR bands (103–106, 138–148, 207–210) were removed from the subimage, yielding 191 bands for the experiments. It was assumed that the image includes five endmembers, i.e., tree, grass, street, roof, and water.

Table IV summaries the unmixing results of all the algorithms. All the methods provide very similar SADs and our method produces slightly better SAD than alternative methods. Fig. 12 presents the resulted endmembers. From the figure, it can be seen that none of the methods gives a good estimate of the grass and street target. This is probably because their spectra are similar to those of trees and water, respectively, which brings difficulty for finding the subspace spanned by the endmembers. The same phenomenon can also be found in the abundance maps shown in Fig. 13. Overall, GMC-NMF achieves competitive unmixing performance against alternatives.

V. CONCLUSION

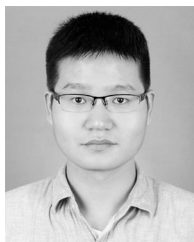
The solution space of NMF-based unmixing is very huge due to its nonconvex objective function and attempting to estimate all variables simultaneously, leading to undesirable unmixing

performance. This issue can be overcome by adding various prior knowledge on abundances, such as sparsity. Following this line, in this article, the GMC penalty is embedded into NMF to limit the number of trivial solutions. Thanks to its capacity of producing unbiased sparsity while preserving the convexity of original NMF with respect to individual variable, the yielded endmembers and abundances are more accurate. Experimental results on both simulated and real-world data confirm the powerfulness and effectiveness of the proposed approach. The method in this article belongs to model-based optimization families, which require tremendous time to converge. In our future work, we will combine the model-based optimization and deep learning methods to develop a model-driven deep hyperspectral unmixing network for time-efficient unmixing.

REFERENCES

- [1] W. Xie, T. Jiang, Y. Li, X. Jia, and J. Lei, "Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4218–4230, Jul. 2019.
- [2] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, Jan. 2020.
- [3] B. Yang, B. Wang, and Z. Wu, "Nonlinear hyperspectral unmixing based on geometric characteristics of bilinear mixture models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 694–714, Feb. 2018.
- [4] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, Jan. 2016.
- [5] R. Heylen, V. Andrejchenko, Z. Zahiri, M. Parente, and P. Scheunders, "Nonlinear hyperspectral unmixing with graphical models," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4844–4856, Jul. 2019.
- [6] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Ann. JPL Airborne Geosci. Workshop*, 1993, pp. 1–14.
- [7] J. M. P. Nascimento and J. M. B. 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.
- [8] D. C. Heinz and Chin-I-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.
- [9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 535–541.
- [10] J. Peng, Y. Zhou, W. Sun, Q. Du, and L. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.2996688](https://doi.org/10.1109/TGRS.2020.2996688).
- [11] R. Huang, X. Li, and L. Zhao, "Spectral-spatial robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8235–8254, Oct. 2019.
- [12] 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, Mar. 2007.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [17] J. Yao, D. Meng, Q. Zhao, W. Cao, and Z. Xu, "Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2991–3006, Jun. 2019.
- [18] W. Wang and Y. Qian, "Adaptive $L_{1/2}$ sparsity-constrained NMF with half-thresholding algorithm for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2618–2631, Jun. 2015.
- [19] Y. Qu and H. Qi, "UDAS: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.
- [20] 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.
- [21] M. Brbi and I. Kopriva, " ℓ_0 -motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.
- [22] J. Abe, M. Yamagishi, and I. Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 4918–4922.
- [23] 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.
- [24] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.
- [25] Z. Zhang, Q. Wang, and Y. Yuan, "Hyperspectral unmixing via $l_{1/4}$ sparsity-constrained multilayer NMF," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2143–2146.
- [26] 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.
- [27] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and ℓ_q sparse regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6371–6384, Nov. 2016.
- [28] I. W. Selesnick and I. Bayram, "Enhanced sparsity by non-separable regularization," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2298–2313, May 2016.
- [29] I. Selesnick and M. Farshchian, "Sparse signal approximation via non-separable regularization," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2561–2575, May 2017.
- [30] A. Marmin, M. Castella, and J. Pesquet, "How to globally solve non-convex optimization problems involving an approximate l_0 penalization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 5601–5605.
- [31] A. Lanza, S. Morigi, I. W. Selesnick, and F. Sgallari, "Sparsity-inducing nonconvex nonseparable regularization for convex image processing," *SIAM J. Imag. Sci.*, vol. 12, no. 2, pp. 1099–1134, 2019.
- [32] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.
- [33] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Nonconvex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7332–7342, Sep. 2018.
- [34] X. Tian, R. Zheng, C. J. Chu, O. H. Bell, L. B. Nicholson, and A. Achim, "Multimodal retinal image registration and fusion based on sparse regularization via a generalized minimax-concave penalty," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 1010–1014.
- [35] L. Yin, A. Parekh, and I. Selesnick, "Stable principal component pursuit via convex analysis," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2595–2607, May 2019.
- [36] N. G. Polson *et al.*, "Proximal algorithms in statistics and machine learning," *Stat. Sci.*, vol. 30, no. 4, pp. 559–581, 2015.
- [37] D. Tuia, R. Flamary, and M. Barlaud, "Nonconvex regularization in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6470–6480, Nov. 2016.
- [38] I. Selesnick, A. Lanza, S. Morigi, and F. Sgallari, "Non-convex total variation regularization for convex denoising of signals," *J. Math. Imag. Vis.*, vol. 62, nos. 6–7, pp. 825–841, 2020.
- [39] 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.
- [40] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [41] 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.
- [42] 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.

- [43] R. N. Clark *et al.*, *USGS digital spectral library splib06a* (No. 231). US Geological Survey, 2007.
- [44] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravorty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, 2019.



Fengchao Xiong (Member, IEEE) received the B.S. degree in software engineering from Shandong University, Jinan, China, in 2014, and the Ph.D. degree from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2019.

He is currently a Lecturer with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He visited Wuhan University as an exchange student from September, 2011 to July, 2012 and Griffith University as a visiting scholar from November, 2017 to

December, 2018. His research interests include hyperspectral image processing, machine learning, and pattern recognition.



Jun Zhou (Senior Member, IEEE) received the B.S. degree in computer science and the B.E. degree in international business from the Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Montreal, Canada, in 2002, and the Ph.D. degree in computer science from the University of Alberta, Edmonton, Canada, in 2006.

He is an Associate Professor with the School of Information and Communication Technology, Griffith University, Nathan, Australia. He was a Research Fellow with the Research School of Computer Science, Australian National University, Canberra, Australia, and a Researcher with the Canberra Research Laboratory, NICTA, Australia. His research interests include pattern recognition, computer vision, and spectral imaging with their applications to remote sensing and environmental informatics.



Jianfeng Lu (Member, IEEE) received the M.S. and Ph.D. degrees in pattern recognition and intelligent system from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.

He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interests include image analysis, machine learning, data mining, and intelligent robot.



Yuntao Qian (Member, IEEE) received the B.E. and M.E. degrees in automatic control from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 1996.

During 1996–1998, he was a Postdoctoral Fellow with the Northwestern Polytechnical University, Xi'an, China. Since 1998, he has been with the College of Computer Science, Zhejiang University, Hangzhou, China, where he became a Professor in 2002. During 1999–2001, 2006, 2010, 2013, 2015–

2016, and 2018, he was a Visiting Professor with Concordia University, Hong Kong Baptist University, Carnegie Mellon University, the Canberra Research Laboratory of NICTA, Macau University, and Griffith University, respectively. His research interests include machine learning, signal and image processing, pattern recognition, and hyperspectral imaging.

Dr. Qian is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.