Hyperspectral Unmixing Based on Nonnegative Matrix Factorization: A Comprehensive Review

Xin-Ru Feng, Heng-Chao Li, *Senior Member, IEEE*, Rui Wang, Qian Du, *Fellow, IEEE*, Xiuping Jia, *Fellow, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Hyperspectral unmixing has been an important technique that estimates a set of endmembers and their corresponding abundances from a hyperspectral image (HSI). Nonnegative matrix factorization (NMF) plays an increasingly significant role in solving this problem. In this article, we present a comprehensive survey of the NMF-based methods proposed for hyperspectral unmixing. Taking the NMF model as a baseline, we show how to improve NMF by utilizing the main properties of HSIs (e.g., spectral, spatial, and structural information). We categorize three important development directions, including constrained NMF, structured NMF, and generalized NMF. Furthermore, several experiments are conducted to illustrate the effectiveness of associated algorithms. Finally, we conclude this article with possible future directions with the purposes of providing guidelines and inspiration to promote the development of hyperspectral unmixing.**

***Index Terms*—Deep learning, hyperspectral unmixing, linear mixture model, nonnegative matrix factorization.**

Nomenclature

HSIs	Hyperspectral images.
DL	Deep learning.
MV	Minimum volume.
PPI	Pixel purity index.
IEA	Iterative error analysis.
VCA	Vertex component analysis.
SGA	Simplex growing algorithm.
ICA	Independent component analysis.
NMF	Nonnegative matrix factorization.
LSTM	Long short-term memory.
LMM	Linear mixture model.
NLMM	Nonlinear mixture model.
ANC	Abundance nonnegative constraint.
ASC	Abundance sum-to-one constraint.
FCLS	Fully constrained least squares.
TV	Total variation.
GPU	Graphics processing units.
PSO	Particle swarm optimization.
BMM	Bilinear mixture model.
MRF	Markov random field.
SLIC	Simple linear iterative clustering.
NTF	Nonnegative tensor factorization.
CPD	Canonical polyadic decomposition.
BTD	Block term decomposition.

Manuscript received January 27, 2022; revised March 29, 2022 and April 26, 2022; accepted May 7, 2022. Date of publication May 16, 2022; date of current version June 10, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871335 and Grant 61801404, and in part by the Fundamental Research Funds for the Central Universities under Grant 2682020ZT35. *(Corresponding author: Heng-Chao Li.)*

Xin-Ru Feng and Rui Wang are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: fengxinru@my.swjtu.edu.cn; wangruinew@163.com).

Heng-Chao Li is with the School of Information Science and Technology, National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: hcli@home.swjtu.edu.cn).

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

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: x.jia@adfa.edu.au).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2022.3175257

I. Introduction

HYPERSPECTRAL images (HSIs) acquired by imaging spectrometers, record hundreds or thousands of spectral bands of the observed scene in a single acquisition [1]–[6]. Owing to the wealthy spectral information, HSIs have been applied to many applications, including agricultural and military defense [7]–[9], food quality control [10], [11], mineralogical mapping of earth surface [12], and pharmaceutical manufacturing industry [13]. Due to the complexity of objects and the relatively low spatial resolution, pixels in HSIs are normally composed of mixed spectral responses from multiple ground objects [4], [14]. Mixed pixels affect the performance of hyperspectral analysis, such as object classification and identification. To address this problem, hyperspectral unmixing is developed to decompose each pixel of an HSI into a set of endmembers and their corresponding abundances.

In general, unmixing algorithms can be divided into four categories: geometrical, sparse regression-based, statistical, and deep learning (DL)-based methods. Geometrical unmixing algorithms work under the assumption that the endmembers of an HSI are the vertices of a simplex with the minimum volume enclosing the dataset or of a simplex with the maximum volume contained in the convex hull of the dataset. Pure pixel-based and minimum volume (MV)-based methods belong to this category. The pure pixel-based algorithms assume that there is one pure pixel at least per endmember. The classical methods include the pixel purity index (PPI) [15], N-FINDR [16], iterative

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

error analysis [17], vertex component analysis (VCA) [18], and the simplex growing algorithm [19], [20]. The MV-based approaches seek a mixing matrix that minimizes the volume of the simplex defined by its columns, such as the minimum volume enclosing simplex [21] and the simplex identification via variable splitting and augmented Lagrangian [22]. To reduce the influence induced by the intrinsic nonlinearity of the geometric manifold of the HSI and extract the endmembers accurately, a novel nonlinear endmember extraction algorithm [23] was proposed by combining the hypergraph framework-based manifold representation and fuzzy assessment. In [24], maximum volume inscribed ellipsoid method was presented to attract endmembers effectively. As one of the mainstream methods, the geometrical unmixing approaches have shown their powerful ability in extracting endmembers from HSIs. However, the geometrical algorithms may hardly extract the endmembers from the highly mixed data since pure spectral signatures are not available.

With the increasing availability of spectral libraries for materials measured on the ground, sparse regression-based methods are proposed by expressing each mixed pixel in a scene as a linear combination of a finite set of pure spectral signatures in a spectral library. Sparse regression-based algorithms avoid estimating the number of endmembers and identifying the endmember signatures in the original dataset [25], [26]. Owing to these two advantages, more efforts have been dedicated to improving the sparse unmixing performance. For example, double weights were introduced in [27] to improve the sparsity of fractional abundances in both spectral and spatial domains, where one is used to enhance the sparsity of endmembers in the spectral library, and the other is to encourage the sparsity of fractional abundances. To make full use of the spatial–contextual information, a new spectral–spatial weighted sparse unmixing framework [28] was developed for hyperspectral unmixing. Besides, the spatial correlation was incorporated to promote the abundance estimation in [29]–[31]. Specifically, a superpixel-based reweighted low rank and total variation [29] method was proposed to enhance the performance of the traditional spatial-regularization-based sparse unmixing approaches. By using multiview collaborative sparse and spectral–spatial weights, the new sparse unmixing model [30] took the advantage of spectral information as well as spatial information. In [31], graph Laplacian regularization was utilized to promote the smoothness of abundance maps in the sparse regression framework. These methods have obtained promising unmixing results. However, the spectra in the library have high coherence and are undesirable due to the diverse imaging conditions, which limit the applicability of these approaches.

The statistical algorithms identify the endmembers and their corresponding abundances at the same time by utilizing the statistical properties of the HSI. Popular statistical algorithms include independent component analysis [32], [33], nonnegative matrix factorization (NMF) [34], [35], and Bayesian approaches [36]–[38]. Among them, NMF provides a good fit for hyperspectral unmixing owing to its nonnegativity and interpretability. Therefore, numerous NMF-based methods have been developed to pursue better unmixing performance.

In the last few years, DL [39], [40] has shown great power and potential in pattern recognition. Therefore, many researchers have focused on hyperspectral unmixing using autoencoder and its variants, achieving more competitive unmixing performance [41]–[47]. In addition, Hong *et al.* [48] proposed an effective guidance for real endmembers with shared weights in the autoencoder-like architecture. In [49], the HSI was processed as sequential data, and a long short-term memory network was included in autoencoder architecture to capture spectral correlation information. These approaches can learn the abundance fractions from the original data via a series of the hierarchical layers, which are more suitable for coping with a variety of situations. For the availability in practice, Deshpande and Bhatt [50] proposed a two-stage fully connected self-supervised DL network for alleviating some practical issues, such as the noise and perturbation. Nevertheless, there are still several drawbacks [51]. For instance, current approaches often require a lot of training samples and network parameters to achieve satisfactory unmixing performance.

Compared with the geometrical methods and sparse regression-based methods, the NMF-based methods are powerful to extract simultaneously the endmembers and their associated abundances. Combining the ability to extract hierarchical features as DL-based approaches, multilayer/deep NMF [52], [53] models have been developed to explore hidden information with interpretability power as in classical NMF.

From the perspective of spectral signatures in HSIs, there are two main challenges. One is spectral variability [54]–[57], which is often brought by many factors, such as changes in illumination, environmental, atmospheric, and temporal conditions. It may lead to large amounts of errors in abundance estimation. To address this problem, a hierarchical sparse NMF (HSNMF) [58] introduced hierarchical sparsity constraints for describing endmember variability. Besides, endmember variability was considered by building a 4-D endmember tensor along with a new low-rank regularization [59] and relying on structured additively tuned linear mixing model [60]. Another issue is the multiple physical interactions in the resulting observed spectrum [14]. This challenge will reduce the generalization ability of unmixing methods based on the linear mixture model (LMM) and increase computational complexity. As such, combined with nonlinear mixture model, NMF can also be applied to form some novel nonlinear unmixing methods [61], [62].

In this article, we aim to provide a survey on NMF-based hyperspectral unmixing. We take the NMF model as a baseline to show how to improve NMF by utilizing the main properties of HSIs (e.g., spectral information, spatial information, and structural information). We introduce three important development directions for the NMF model and discuss their pros and cons.

1) *Constrained NMF:* By introducing additional constraints or penalty terms to the cost function, such as sparsity constraints, smooth constraints, and graph constraints.
2) *Structured NMF:* By modifying the structure of the cost function, e.g., weighted NMF, convex NMF, robust NMF.
3) *Generalized NMF:* By extending the decomposition form, involving nonnegative tensor factorization (NTF), multilayer NMF, deep NMF, etc.

In addition, we conduct several experiments to demonstrate the effectiveness of some associated algorithms. The purpose is

to give guidelines and inspiration for the future improvement of hyperspectral unmixing.

The rest of this article is organized as follows. In Section II, we give a general introduction to spectral mixture model and the unmixing problem. Sections III, IV, and V review classical NMF unmixing categories according to different constraints, structures of the cost function, and decomposition forms, respectively. Extensive experiments are conducted and the results are discussed in Section VI. Section VII draws comprehensive conclusions and presents a brief outlook on future possible research directions.

II. Classical NMF for Hyperspectral Unmixing

Let $\mathcal{X} \in \mathbb{R}^{r\times c\times B}$ denote an HSI with B bands, r rows, and c columns. Through the unfolding operation, the HSI can be represented by a matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1P} \\ \vdots & \ddots & \vdots \\ x_{B1} & \cdots & x_{BP} \end{bmatrix} \in \mathbb{R}^{B\times P} \tag{1}$$

with the number of pixels $P = r \times c$. The element at (b, p) denoted by x_{bp} represents the reflection value from bth band of pth pixel. From the row perspective, $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^B]$, where the bth vector $\mathbf{x}^b$ is the ground information in the bth band. Generally, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]$ is given from the column perspective, where the pth vector $\mathbf{x}_p$ is the spectrum of pth pixel. The LMM assumes that an observed pixel spectrum in an HSI can be produced by a linear combination of endmember signatures and their corresponding abundances. The matrix formulation of the LMM can be described as

$$\mathbf{X} = \mathbf{AS} + \mathbf{G} \tag{2}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_M] \in \mathbb{R}^{B\times M}$ denotes endmember matrix, the vector $\mathbf{a}_m$ represents the mth endmember signature, and M denotes the number of endmembers. $\mathbf{S} \in \mathbb{R}^{M\times P}$ represents the abundance matrix for all endmembers, and $\mathbf{G} \in \mathbb{R}^{B\times P}$ represents the noise matrix. Typically, two constraints are imposed on $\mathbf{S}$, i.e., the abundance nonnegative constraint and the abundance sum-to-one constraint (ASC), given by $\mathbf{S} \geq 0$ and $\mathbf{1}_M^T\mathbf{S} = \mathbf{1}_P^T$, where $\mathbf{1}_M$ and $\mathbf{1}_P$ are all-one column vectors with size M and size P, respectively, and $(\cdot)^T$ denotes the transpose operation.

Given a matrix $\mathbf{X}$, NMF [63] focuses on decomposing it into the product of two nonnegative matrices $\mathbf{A}$ and $\mathbf{S}$, i.e., $\mathbf{X} \approx \mathbf{AS}$. Obviously, this decomposition form is consistent with LMM. Thus, NMF is attractive for hyperspectral unmixing. In general, Frobenius norm is utilized to measure the approximation between $\mathbf{X}$ and $\mathbf{AS}$, and the cost function is expressed as

$$\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|_F^2, \text{s.t.} \mathbf{A} \geq 0, \mathbf{S} \geq 0, \mathbf{1}_M^T\mathbf{S} = \mathbf{1}_P^T \tag{3}$$

where the operator $\|\cdot\|_F$ denotes the Frobenius norm. The multiplicative update rules are deduced as

$$\mathbf{A} \leftarrow \mathbf{A} \odot (\mathbf{XS}^T) \oslash (\mathbf{ASS}^T) \tag{4a}$$

$$\mathbf{S} \leftarrow \mathbf{S} \odot (\mathbf{A}^T\mathbf{X}) \oslash (\mathbf{A}^T\mathbf{AS}) \tag{4b}$$

in which $\odot$ and $\oslash$ stand for the elementwise multiplication and division, respectively. Meanwhile, when abundance matrix $\mathbf{S}$ is updated, the ASC requires to be satisfied by redefining the observation and spectral signature matrices as

$$\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta\mathbf{1}_P^T \end{bmatrix}, \quad \bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \delta\mathbf{1}_M^T \end{bmatrix} \tag{5}$$

where δ is a parameter to control the impact of the ASC. For the implementation of NMF, a crucial issue is how to initialize the related variables. The endmember matrix $\mathbf{A}$ can be initialized by the various methods, such as random values from 0 to 1, VCA [18], and automatic target generation process [64]. Abundance matrix $\mathbf{S}$ is initialized according to the fully constrained least squares algorithm [65]. To speedup the NMF processing, an adaptive projected NMF algorithm was parallelized by introducing its parallel version in [66]. In addition, by applying Nesterov's optimal gradient method, NeNMF [67] was proposed to accelerate the optimization.

Nevertheless, the cost function of NMF is nonconvex so that it easily falls into local optimal solutions. As such, to improve the unmixing performance, there are three important improvement directions for the NMF model. A considerable number of NMF-based methods address the spectral and spatial information by introducing additional constraints or penalty terms to the cost function. These methods are reviewed in Section III. As discussed in Section IV, many methods enable flexibility to account for more structures and details, such as the difference of pixels, bands, and elements. Moreover, a lot of methods extend the decomposition form to acquire more essential characteristics, e.g., nonlinearities, 3-D structure information, hidden information. Such methods are considered in Section V. The framework of the NMF-based methods for hyperspectral unmixing is illustrated in Fig. 1. The specifical unmixing methods are mainly summarized in Table I.

III. Constrained NMF for Hyperspectral Unmixing

By exploiting the spectral and spatial information in HSIs, additional constraints have been imposed on the endmembers and abundances to obtain better unmixing performance. The constrained NMF model can be integrated as

$$\begin{aligned} &\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \alpha J(\mathbf{A}) + \beta J(\mathbf{S}) \\ &\text{s.t. } \mathbf{A} \geq 0, \mathbf{S} \geq 0, \mathbf{1}_M^T\mathbf{S} = \mathbf{1}_P^T \end{aligned} \tag{6}$$

where $J(\mathbf{A})$ and $J(\mathbf{S})$ are regularization terms for endmembers and abundances, respectively, and α and β are nonnegative parameters to balance the effect of the corresponding constraint terms. Next, we mainly describe the algorithms that incorporate the constraints on endmember matrix in Section III-A. By contrast, numerous works are reported in terms of imposing constraints for abundances, given in Section III-B to III-G. More details are presented as follows.

A. Endmember Constraints

The constraints for endmembers are integrated into NMF by minimizing simplex volume [68]–[72], compacting endmember

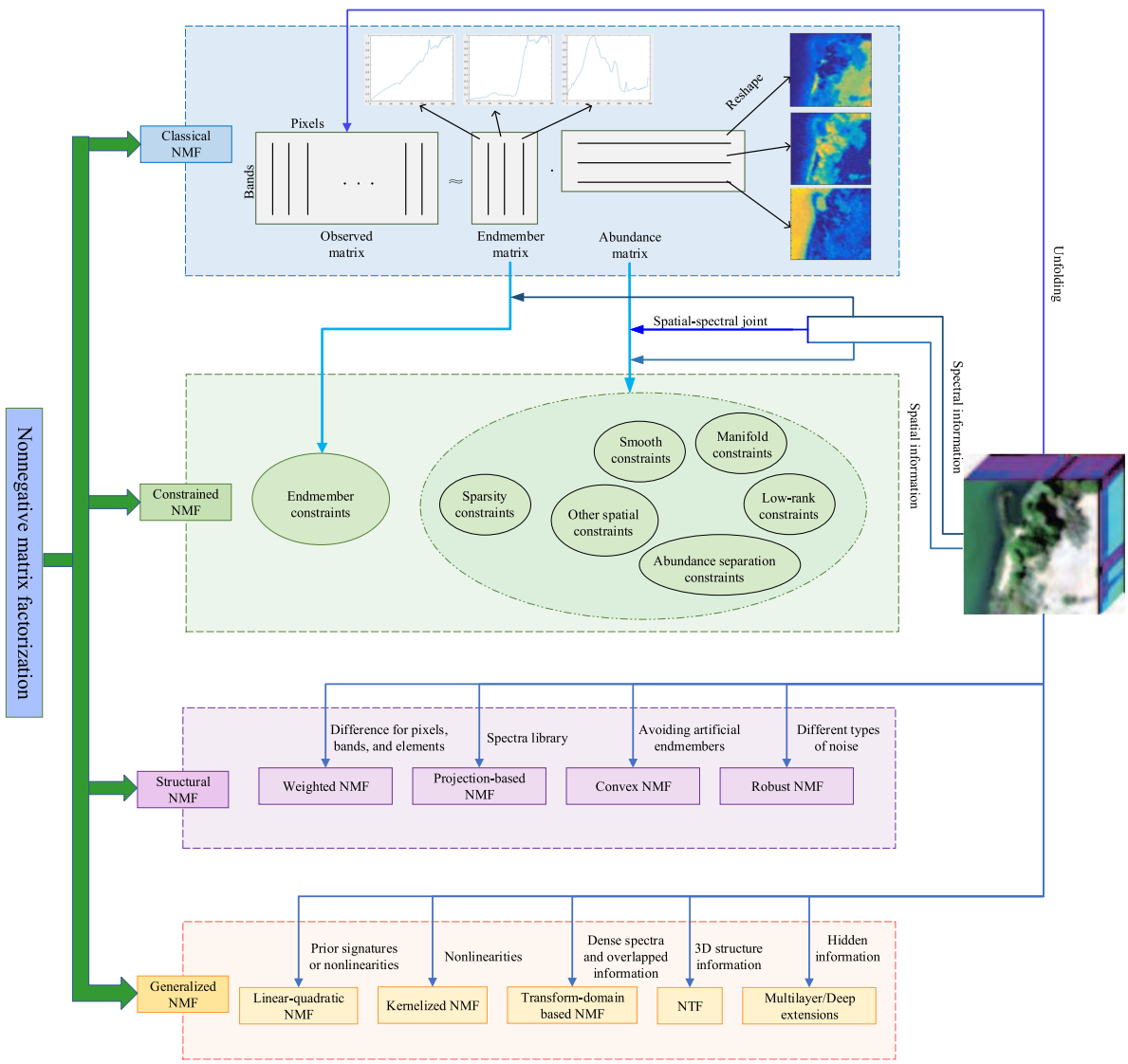


Fig. 1. Framework of the NMF-based methods for hyperspectral unmixing.

distance [73]–[77], keeping signature smoothness [78]–[81], introducing prior spectral information [82], and exploring high-level semantic information [83].

Motivated by geometric insights, the minimum volume constraint (MVC) is robust without the pure pixel assumption, which can be presented as

$$J(\mathbf{A}) = Vol(\mathbf{A}) \tag{7}$$

where $Vol(\mathbf{A})$ is the volume of the simplex whose vertices correspond to the endmembers $\mathbf{A}$. As a typical one, the first minimum-volume NMF was proposed in [68] by incorporating the MVC into the NMF to effectively extract endmembers from highly mixed data. It should be noted that $J(\mathbf{A})$ is calculated with dimensionality reduction since the matrix determinant is valid only for square one. On this basis, Qu and Bao [69] further introduced total variation (TV) and reweighted sparse regularizers to form a multiple-priors ensemble constrained NMF (MPEC-NMF) method. Although the above-mentioned methods do not require the pure-pixel assumption, it is inconvenient to adopt appropriate dimensionality reduction. Therefore, more algorithms were proposed to define the volume of the simplex. In [70], three different volume regularizers were presented to form volume-regularized NMF (VRNMF), including Gram determinant, logarithm Gram determinant, and nuclear norm. Besides, volume minimization can be promoted by pushing endmembers toward a solution that is quadratically regularized by a given simplex [71], [72].

Different from using simplex volume as an endmember constraint, Yu *et al.* [73] used endmember distance (EMD) to keep the simplex as compact as possible, and proposed a minimum distance constrained NMF (MDC-NMF) method. The EMD is defined as the sum of distances from each endmember to their centroid, expressed as

$$J(\mathbf{A}) = \sum_{m=1}^{M} \|\mathbf{a}_m - \frac{1}{M}\sum_{m=1}^{M} \mathbf{a}_m\|_2^2. \tag{8}$$

TABLE I
TAXONOMY OF THE NMF-BASED METHODS FOR HYPERSPECTRAL UNMIXING

CATEGORY	SUBCATEGORY	TYPICAL EXAMPLES
Constrained NMF	Endmember constraints	MVC-NMF [68], **MPEC-NMF** [69], VRNMF [70], R-CoNMF [71], **ICoNMF-TV** [72], MDCNMF [73], CSNMF-GPU [74], APSO and MOPSO [75], BCNMF [76], IP-NMF [77], PSnsNMF and PSNMFSC [78], MiniDisCo [79], EDCNMF [80], Arctan-NMF [81], **MRS-PPK** [82], MCG-NMF [83].
	Sparsity constraints	PSnsNMF and PSNMFSC [78], **$L_{1/2}$-NMF** [84], **HT$L_{1/2}$-NMF** [85], DgS-NMF [86], **DGC-NMF** [87], NMF-SMC [88], CSNMF-GPU [74], Arctan-NMF [81], TV-RSNMF [89], RNLCF [90], DSPLR-NMF [91], GMCA-HU [92], TS-NMF [93], CSsRS-NMF [94], GMC-NMF [95], R-CoNMF [71], Sparsity constrained distributed unmixing [96], SGSNMF [97], NLNMF [98].
	Smooth constraints	PSnsNMF and PSNMFSC [78], ASSNMF [99], WNMF [100], DAC^{2}NMF [101], Sparsity constrained distributed unmixing [96], TV-RSNMF [89], **ICoNMF-TV** [72], **MPEC-NMF** [69], NLHTV-LSRNMF [102].
	Manifold constraints	**GLNMF** [103], **MRS-PPK** [82], **SC-GLR$_{l_1}$NMF** [104], HGNMF [105], GNMF [106], **SS-NMF** [107], **BF-L_2SNMF** [108], GLrNMF [109], **HG$L_{1/2}$-NMF** [110], **R-NMF** [111], **SRASU** [112], SS-NMF [113], **SC-NMF** [114].
	Low-rank constraints	DSPLR-NMF [91], GLrNMF [109].
	Abundance separation constraints	ASSNMF [99], DAC^{2}NMF [101], Semi-supervised NMF [115].
	Other spatial constraints	**CSNMF** [116], **KLSNMF** [117], **Subspace structure regularized sparse NMF** [118], **SDSNMF** [119], S^2-NMF [120], Semi-supervised NMF [115].
Structured NMF	Weighted NMF	CW-NMF [121], SpNMF [122].
	Projection-based NMF	**PNMF** [123].
	Convex NMF	RCMF [124], SC-NMF [114], HSNMF [58].
	Robust NMF	ISB-NMF$_{L_0}$ [125], rNMF [62], **$L_{1/2}$-RNMF** [126], SSRNMF [127], SS-NMF [113], CENMF [128], CSsRS-NMF [94], RNLCF [90], CANMF-TV [129], MCSNMF [130], GLNMF [131].
Generalized NMF	Linear-quadratic NMF	NMFupk [132], Method proposed in [133], Fan-NMF [134], Semi-NMF [61], RASU [135].
	Kernelized NMF	CKNMF [136], biobjective NMF [137], OKNMF [138], IKNMF and IIKNMF [34].
	Transform-domain based NMF	Method proposed in [139], TS-NMF [93].
	NTF	Method proposed in [140], MV-NTF [141], S-MV-NTF [142], MV-NTF-TV [143], Method proposed in [144], DWSNTF [145], EIC-NTF [146], SPLRTF [147], **SCNMTF** [148], ULTRA [149], ULTRA-V [59], BUTTDL1 [150], Method proposed in [151].
	Multilayer/deep extensions	**MLNMF** [52], DCMLNMF [152], **SSTV-MLNMF** [153], **AGMLNMF** [154], **HR-MLNMF** [155], HP-MLNMF [35], **CMLNMF** [156], 2LFKAA [157], AC-MLKNMF [158], **SDNMF-TV** [53], L_1-DNMF [159], SSRDMF [160], MNN-BU [161], SNMF-Net [162].

Boldfaced letters represent that $L_{1/2}$ regularizer is employed for the corresponding methods.

This constraint is not only simpler but also convex for endmember matrix $\mathbf{A}$. Under the framework of MDC-NMF, Wu *et al.* [74] reported an unmixing method along with sparsity constraint and graphics processing units (GPU). Yang *et al.* [75] introduced particle swarm optimization (PSO) to solve the optimization problem, which possessed good global search ability and convergence. Afterward, the bilinear mixture model (BMM)-based constrained NMF algorithm (BCNMF) was presented in [76] with the EMD constraint for unsupervised nonlinear spectral unmixing, in which pixels were projected into their approximate linear mixture components based on the characteristics of BMM to reduce the collinearity greatly. Similarly, in [77], an inertia constraint was presented so as to promote the homogeneity of estimated spectra from the same class using the trace of the covariance matrix. It can deal with intraclass variability by extracting a separate set of pure material spectra from each observed pixel spectrum.

To promote the smoothness, different metrics are applied on the matrix $\mathbf{A}$. Specifically, an adaptive potential function from discontinuity adaptive Markov random field (MRF) model was adopted in [78]. Spectral dispersion function [79] encouraged the variance of each endmember spectrum to be as low as possible. In [80], an endmember dissimilarity function has been defined to make the estimated endmember signatures to be smooth. Besides, a quadratic weighted norm was used as a regularizer term in [81] to exploit spectral smoothness.

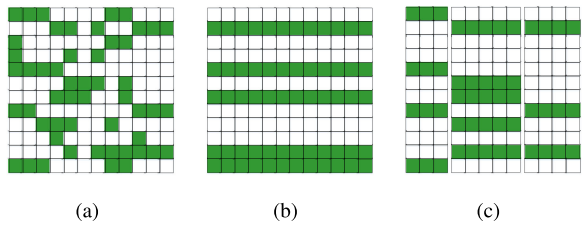


Fig. 2. Illustration of (a) sparsity, (b) collaborative sparsity, and (c) group sparsity.

By incorporating prior spectral information, Tong *et al.* [82] introduced a constraint to minimize the differences between the estimated endmember and the known one. It not only considers the discrepancies between the standard signature and the corresponding one in the image, but also shows flexibility in terms of its extensions.

Recently, in order to effectively learn high-level semantic information, a multiple clustering guided NMF (MCG-NMF) unmixing approach was proposed in [83]. Specifically, clustering analysis via the k-means method is conducted on the data matrix $\mathbf{X}$. Then, selecting K pixels in each cluster, according to the principle of k-nearest neighbors. Finally, the $\mathbf{R}^k, k = 1, \ldots, K+1$ is constructed. Thus, the regularization term is expressed as

$$J(\mathbf{A}) = \frac{1}{K+1} \sum_{k=1}^{K+1} \|(\mathbf{R}^k - \mathbf{A})\mathbf{W}\|_F^2 \tag{9}$$

where $\mathbf{W}$ is a diagonal matrix and depends on the similarity relationship between each pixel and their cluster mean.

B. Sparsity Constraints

Sparsity regularizer is extensively exploited during the hyperspectral unmixing procedure since the distribution of each endmember is sparse in general. Inspired by [97], we discuss the works on imposing sparsity regularizer from sparsity, collaborative sparsity, and group sparsity perspectives. As shown in Fig. 2, diverse effects can be promoted for the abundances, where zero coefficients are denoted by white squares.

1) Sparsity: Jia and Qian [78] proposed two algorithms (called PSnsNMF and PSNMFSC) to explicitly represent sparsity. In particular, PSnsNMF controls the sparsity using the parameter of smoothing matrix in nonsmooth NMF (nsNMF) [163]. Thus, it is difficult to balance the smoothness and sparsity. PSNMFSC enforces sparsity by setting both L_1-norm and L_2-norm. However, the algorithm needs to assign an exact sparsity level which cannot be known a priori, and sparsity levels of different endmembers also vary from each other. In order to encourage the sparsity of abundance matrix, L-norm is widely used in the field of unmixing. Along this line, L_0 regularizer counts the number of non-zero elements in the abundances to yield the sparsest results, while its optimization belongs to an NP-hard problem. Therefore, L_q regularizer is attractive in real applications, expressed as

$$J(\mathbf{S}) = \|\mathbf{S}\|_q = \sum_{m,p=1}^{M,P} (s_{mp})^q \tag{10}$$

where s_{mp} represents the element of the matrix $\mathbf{S}$ in the mth row and pth column.

1) By setting $q = 1$, L_1 regularizer is a popular alternative for achieving a sparse abundance matrix [25], [164]. Nevertheless, it is not compatible with the ASC during the unmixing procedure.
2) $L_q(0 < q < 1)$ regularizer achieves sparser results compared with the L_1 counterpart. Especially, Qian *et al.* [84] showed that $q = 1/2$ was an optimal choice and proposed an $L_{1/2}$-NMF algorithm for unmixing the hyperspectral data. Due to the simplicity and effectiveness, the $L_{1/2}$ regularizer is widely applied to develop novel unmixing approaches (see the boldfaced methods in Table I). To solve the nonconvex optimization caused by the $L_{1/2}$ regularizer, a fast and efficient adaptive half-thresholding algorithm was proposed in [85]. Considering that the mixed level of each pixel may be different from each other, a data-guided sparsity was provided to adaptively employ $L_q(0 < q < 1)$ constraint [86]. In detail, a data-guided map (DgMap) was first learnt by measuring the uniformity of neighboring pixels, thus obtaining adaptive value for q.
3) The problem (15) with $q = 2$ is often considered to improve the unmixing performance as well. For instance, Huang *et al.* [87] proposed a data-guided constrained NMF (DGC-NMF) model by imposing sparsity with either $L_{1/2}$-norm or L_2-norm. Specifically, the sparsity of each pixel is measured first. Then, the $L_{1/2}$ regularizer is utilized to constrain the abundances of the pixels with a high sparsity level, whereas the L_2 regularizer is adopted for generating smooth results with a low sparsity level.

Nevertheless, an $L_q(0 < q < 1)$ regularizer brings challenges as well since it is noncontinuous and nondifferentiable. Accordingly, more efforts are reported to achieve better abundances. Specifically, by adopting a higher order norm, NMF with S-measure constraint (NMF-SMC) was constructed [88]. Since NMF-SMC is a gradient-based algorithm, the convergence speed may be slow for large scale dataset. Wu *et al.* [74], defined the sparsity constraint for the abundances using the difference between L_1-norm and L_2-norm. Furthermore, $\arctan$ function [81] was introduced to explore the sparse property of the abundance matrix.

In order to pursue sparser representation, He *et al.* [89] proposed to utilize a weighted sparse regularizer for abundances, expressed as

$$J(\mathbf{S}) = \|\mathbf{W} \odot \mathbf{S}\|_1 \tag{11}$$

where $\mathbf{W}$ is a weight matrix whose element at (m, p) is calculated by $w_{mp} = 1/(|s_{mp}| + \varepsilon)$, and ε is a positive parameter. It encourages the sparsity of the abundance matrix from the column perspective (i.e., spectral domain). Similarly, a double reweighted L_1-norm regularizer is designed to further exploit the sparsity of abundances in spatial domains [27], [28], [145].

Moreover, to ensure the sparsity of the learned coefficients, a local coordinate constraint was imposed to develop a robust nonnegative local coordinate factorization (RNLCF) method along with a correntropy induced metric (CIM) [90], where the weight was defined as $w_{mp} = \|\mathbf{a}_m - \mathbf{x}_p\|^2$. The generalization in terms of an L_q-norm was presented in [96]. By further accounting for local spatial information, the L_1-norm was added to each nonoverlapping subblock [91].

Based on generalized morphological component analysis (GMCA), each $\mathbf{s}^m$ (the vector of $\mathbf{S}$ in mth row, $m = 1, \ldots, M$) can be modeled as a linear combination of multiple morphological components (s_{mk}) that can be sparsely represented by its associated dictionary basis, i.e., $\mathbf{s}^m \approx \sum_k s_{mk} = \sum_k \mathbf{D}_k \varphi_{mk}$ [92]. A concatenated dictionary $\mathbf{D}$ acts as a discriminator between different morphological components, and φ_{mk} is the sparse representation coefficient constrained by the L_1-norm. Furthermore, in order to simplify the solution, the sparse constraint $\|\varphi\|_1$ is transferred as

$$J(\mathbf{S}) = \sum_{m=1}^{M} \|\mathbf{s}^m\|_1. \tag{12}$$

Through GMCA, spatial information can be naturally considered in the unmixing process by exploiting the sparsity and morphological diversity of the abundance maps associated with each endmember.

Unlike imposing a sparsity regularizer on abundances directly, a transform domain-based sparse regularizer was proposed in [93], expressed as

$$J(\mathbf{S}) = \|\mathbf{S}W^T\|_1 \tag{13}$$

where W is the curvelet transform basis. Very recently, correntropy-based adaptive sparsity constraint [94] was imposed to abundances for each pixel. Besides, a generalized minimax concave (GMC) sparsity regularizer was embedded into NMF [95], which is nonconvex and nonseparable, avoiding systematic underestimation of high components of sparse vector and producing more accurate sparse approximation.

2) Collaborative Sparsity: As shown in Fig. 2(b), collaborative sparsity enforces the row sparsity (joint sparsity), whose formulation is

$$J(\mathbf{S}) = \|\mathbf{S}\|_{2,1} = \sum_{m=1}^{M} \|\mathbf{s}^m\|_2. \tag{14}$$

Since it is challenging to correctly identify the number of endmembers, an overestimation for it was studied in [71] and a collaborative sparsity regularizer is imposed to remove the redundant endmembers. Through this regularizer, the sparsity among the endmembers is achieved simultaneously (collaboratively) for all pixels, i.e., collaborative sparsity of the abundance matrix. In addition, a weighted $L_{2,1}$-norm regularizer was applied to local similar abundances (i.e., blocks) to consider both sparsity and spatial information [165].

3) Group sparsity: As shown in Fig. 2(c), by fully exploring the spatial group structure and sparsity of the HSIs, spatial group sparsity regularizer was proposed in [97], defined as

$$J(\mathbf{S}) = \sum_{g=1}^{G} \sum_{\mathbf{s}_p \in \vartheta_g} c_p \|\mathbf{W}^g \mathbf{s}_p\|_2 \tag{15}$$

where $\mathbf{s}_p (p = 1, \ldots, P)$ denotes the vector of $\mathbf{S}$ in the pth column, and G is the number of the spatial groups (i.e., superpixels), which was generated by an improved simple linear iterative clustering (SLIC) algorithm. That is, the abundance matrix is divided into G groups as $\mathbf{s}_r = (\bar{\mathbf{s}}^1, \ldots, \bar{\mathbf{s}}^G) \in \mathbb{R}^{M \times P}$ in which $\bar{\mathbf{s}}^g$ denotes spatial group ϑ_g. $c_p = \frac{1}{D_p^g}$ (D_p^g is spatial-spectral distance) is a pixelwise confidence index for relaxing the group sparsity constraints of heterogeneous pixels, such as boundaries and small targets. $\mathbf{W}^g$ is updated iteratively to appropriately control the nonzero abundances of ϑ_g. This regularizer is expected to promote the same sparse structure for pixels within a local spatial group. By further combining with nonlocal spatial information, Yang *et al.* [98] developed NLNMF to address the unmixing problem. Compared with the unmixing methods with smooth constraints, it is more reasonable and effective to utilize this local spatial groups with irregular shapes. Furthermore, group sparsity [166] was investigated in unmixing process that accounts for spectral variability through the use of group two-level mixed norm, i.e., $L_{G,p,q} = \|\mathbf{S}\|_{G,p,q}$. More concretely, the fractional LASSO with $L_{G,1,q}, 0 < q < 1$ aims to simultaneously enforce group and within-group sparsity.

As discussed previously, the L_q regularizers draw much attention to enforce the sparsity of abundances, especially for L_1-norm and $L_{1/2}$-norm. Compared with the L_1-norm, the $L_{1/2}$-norm [84] is excellent to obtain a sparser solution, thus enhancing the unmixing performance. Nevertheless, it requires to be combined with other constraints (e.g., graph constraint [103]) such that the intrinsic structures of HSIs can be considered. In addition, many approaches have been devoted to incorporate the spectral and spatial information into the L_1-norm. For example, the reweighted L_1-norm regularizer [89] promotes the sparsity of the abundance matrix from spectral domain, and double reweighted L_1-norm regularizer [145] aims to further describe the sparsity in spatial domains. In this way, the improvement can be obtained greatly in the unmixing process. However, they are sensitive to noise corruption. Collaborative sparsity [71] is helpful to induce the row sparsity, whereas it is incompetent to explore the spatial information. Thus, Huang *et al.* [165] applied a weighted $L_{2,1}$-norm regularizer on blocks and imposed TV regularizer. Meanwhile, it may be more reasonable to encourage the collaborative sparsity when the endmember matrix is composed of spectral library or bundles. Moreover, the group sparsity regularizer considers sparsity at the group level by integrating the spatial group structure [97], [98], but how to effectively group the abundances deserves more investigation.

C. Smooth Constraints

Neighboring pixels are more likely to be constituted by the same materials. Accordingly, it is significant to investigate the spatial correlation among the neighboring pixels. Smooth

constraints are often used to preserve the spatial-contextual information.

An adaptive potential function from discontinuity adaptive MRF model [78] was adopted to promote the piecewise smoothness of abundances, given as

$$J(\mathbf{S}) = g(\mathbf{S} - \mathbf{S}_{\mathcal{N}}) \tag{16}$$

where $\mathbf{S}_{\mathcal{N}}$ is the neighborhood matrix of $\mathbf{S}$. However, PSNMFSC cannot perform well for many real datasets and has high computational complexities [99]. Hence, Liu *et al.* [99] proposed abundance separation and smoothness constrained NMF (ASSNMF). They defined the smoothness function as

$$J(\mathbf{S}) = \frac{1}{2} \times \frac{1}{8} \sum_{h=1}^{8} \|\tilde{\underline{\mathbf{S}}}_h^m - \underline{\mathbf{S}}^m\| \tag{17}$$

where $\tilde{\underline{\mathbf{S}}}_h^m$ is a weight matrix corresponding to $\underline{\mathbf{S}}^m \in \mathbb{R}^{r \times c}$, which is obtained via the reshape operation to the mth row of $\mathbf{S}$. Compared with PSNMFSC where only adjacent elements are considered, ASSNMF is more effective by exploiting its surrounding elements, and is acquired by a linear transform with a low computational cost. Nevertheless, it is not always true that the smoothness levels of two-pixel pairs are the same even if the spatial distances between them are the same.

To describe spatial correlations, a weighted nonnegative matrix factorization (WNMF) algorithm was presented in [100], and the designed weight is expressed as

$$J(\mathbf{S}) = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{(i)}} w_{ij} \|\mathbf{s}_i - \mathbf{s}_j\|_2^2 \tag{18}$$

where w_{ij} is a weight to characterize how much the neighboring pixel $\mathbf{x}_j$ contributes to the considered pixel $\mathbf{x}_i$. This regularizer integrates both spectral and spatial information. Similarly, an adaptive local neighborhood weight constraint was designed to propose a double abundance characteristics constrained NMF (DAC^2 NMF) [101] along with a separation constraint to prevent oversmoothness. In [96], the weight constraint with the L_2-norm was generalized to use the L_q-norm.

Recently, the abundance maps are assumed to be piecewise smooth. Therefore, the TV regularizer is accounted for capturing the piecewise smoothness structure of each abundance map to enhance the unmixing results. He *et al.* [89] first embedded the TV regularizer into the NMF framework, expressed as

$$J(\mathbf{S}) = \|\mathbf{S}\|_{\text{HTV}} = \sum_{m=1}^{M} \|\mathcal{F}\mathbf{s}^m\|_{\text{TV}} \tag{19}$$

where $\mathcal{F}$ denotes the reshape operation from a vector with P pixels to a matrix with the size of $c \times r$, and $\|\cdot\|_{\text{TV}}$ is the anisotropic TV norm. Together with collaborative sparsity and endmember constraint, Yuan *et al.* [72] presented an improved collaborative NMF and TV algorithm (ICoNMF-TV) for the unmixing task. Similarly, multiple-priors ensemble constrained NMF (MPEC-NMF) [69] integrated the MVC, reweighted $L_{1/2}$-norm, and TV regularizers. Besides, Yao *et al.* [102] extended the piecewise smoothness to the nonlocal smoothness, developing a nonlocal TV and log-sum regularized NMF (NLTV-LSRNMF) method.

For the above-mentioned methods, the neighborhood is determined by a set of pixels involved in a predefined regular shape, such as a cross or square window. Consequently, smooth constraints can be utilized to exploit spatial-contextual information adequately, while ignoring the edges and local spatial details. By comparison, it may be more reasonable to use irregular shapes adaptively.

D. Manifold Constraints

The above-mentioned methods exploit the Euclidean structure of the hyperspectral data space. Considering that the data are more likely to lie on a low-dimensional submanifold embedded in the high-dimensional ambient space [167], the intrinsic manifold structure receives increasing attention in the hyperspectral unmixing field. The graph regularizer can be expressed as

$$J(\mathbf{S}) = \sum_{i,j=1}^{P} \|\mathbf{s}_i - \mathbf{s}_j\|^2 \mathbf{W}_{ij} = \text{Tr}\left(\mathbf{S}\mathbf{L}\mathbf{S}^T\right) \tag{20}$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, $\mathbf{W}$ is an affinity matrix, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, and $\mathbf{D}$ is a diagonal matrix with its (i, i) element calculated by $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$. Next, we introduce how to construct the affinity matrix.

1) Spectral Similarity: In [103], a manifold regularizer was incorporated into the sparsity constrained NMF, presenting graph-regularized $L_{1/2}$-NMF (GLNMF) for hyperspectral unmixing, where the affinity matrix of the nearest neighbor graph $\mathbf{W} \in \mathbb{R}^{N \times N}$ is built to model the local structural information, given as

$$\mathbf{W}_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}\right) \tag{21}$$

which is known as the heat kernel. Here, $\mathbf{x}_i$ is one of the k-nearest neighbors of $\mathbf{x}_j$. When $\mathbf{x}_i$ and $\mathbf{x}_j$ are similar, $\mathbf{W}_{ij}$ is relatively large, and σ denotes the standard deviation. This regularizer was leveraged in [82] along with partially known endmembers for unmixing. Inspired by the graph regularizer based on L_2-Laplacian, Rathnayake *et al.* [104] developed an L_1-norm-based graph, called GLR$_{l_1}$. Besides, a Hessian graph [105] was adopted for hyperspectral unmixing. Compared with Laplacian graph, it is more stable and further captures the relationship of the nonlinear mapping.

2) Spectral–Spatial Similarity: By respectively defining spatial geometric distance and spectral geometric distance of these two pixels in local window, dual Laplacian manifold regularizer [106] was established to exploit the geometric structure of the HSI. In order to encode the manifold structures embedded in the hyperspectral data space, a graph Laplacian is incorporated from spatial distance and feature distance perspectives simultaneously [107], where the weight matrix is obtained by using spectral angle distance (SAD) metric and two conditions are adopted to determine the neighbors: one is the nearest spatial distance via a local window, and the other is the nearest feature distance via the SAD similarities in the local window. In this way,

the edges can be preserved by avoiding the graph across dissimilar pixels. Likewise, a bilateral filter regularizer based on graph theory was adopted to utilize the correlation information in both spatial and spectral spaces [108]. Moreover, GLrNMF [109] defined a spatial–spectral semantic weight based on the intersection (denoted by QG) of spatial neighbor (denoted by Q) and spectral neighbor (denoted by G). Motivated by hypergraph learning, the spectral–spatial joint structure was modeled by a hypergraph to capture the high-order relation among multiple pixels [110].

3) Region-Based Similarity: Considering that the spectra are similar in the same region while different between regions, Tong *et al.* [111] presented a region-based structure preserving NMF (R-NMF) to explore consistent data distribution in the same region, aiming at discriminating different data structures across regions. Furthermore, in view of the difference in spatial structure, an HSI was partitioned into homogeneous region and detail region based on a sketch map [112], in which the manifold constraint and the $L_{1/2}$ regularizer were employed for the homogeneous region, while the L_1 regularizer for detail region.

4) Other weight learning

Different from the above-mentioned perspectives to establish a graph regularizer, the work in [113] was based on the local linear embedding assumption, where the weight matrix $\mathbf{W}$ is learned by minimizing the following equation:

$$\mathbf{W}_{ij} = \arg\min \|\mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \mathbf{W}_{ij}\mathbf{x}_j\|_2 \tag{22}$$

to exploit the local geometric structure. In addition, the clustering algorithm is also beneficial to characterizing the structure information of the HSIs. In [114], subspace clustering was applied to capture the latent characteristic structure, for which the weight matrix $\mathbf{W}$ is constructed by

$$\mathbf{W} = \frac{\mathbf{H} + \mathbf{H}^T}{2} \tag{23}$$

to form a graph regularizer. Here, $\mathbf{H}$ is the subspace structure matrix that is learned by minimizing $\|\mathbf{X} - \mathbf{XH}\|_F^2$, s.t. $\mathrm{diag}(\mathbf{H}) = 0$. It is noteworthy that only the largest k values are remained for each column of $\mathbf{H}$.

E. Low-Rank Constraints

The high spatial correlation of HSIs can be also translated into the low rank of the involved abundance matrices.

Let G denote the number of non-overlapping subimages from the input HSI. By enforcing simultaneously the local low-rank and sparse structures to the abundance matrix $\mathbf{S}_G$, DSPLR-NMF [91] was reported for unmixing, in which, the local low-rank constraint is given as

$$J(\mathbf{S}) = \sum_{g=1}^{G} \|\mathbf{S}_g\|_* \tag{24}$$

where $\|\cdot\|_*$ denotes the nuclear norm. In order to naturally incorporate spatial priors, superpixels were generated by employing SLIC algorithm to the HSI and constrained using the low-rank penalty [109]. For the learning of subspace structure [118], the low-rank constraint was utilized to construct a self-representation matrix.

F. Abundance Separation Constraints

Taking the spatial correlation into consideration, the unmixing results can be improved and robust to noise generally. However, the oversmoothing problem may occur since dissimilar pixels are often ignored in a local window. Therefore, an abundance separation constraint [99] was imposed based on the KL divergence to minimize the correlation between different endmembers. Likewise, Liu *et al.* [101] introduced the separation constraint to preserve the inner diversity of the same type of materials. Moreover, in [115], a dissimilarity regularizer constructed by label information was incorporated into the NMF.

As discussed previously, numerous works are devoted to obtaining better spatial structure from different insights, such as sparsity, spatial–contextual information, low-dimensional manifold structure, low-rank structure, and global spatial information. Among these constraints, it has considerable potential to make use of the spatial and spectral information simultaneously for HSIs applications.

G. Other Spatial Constraints

To maintain the structural information, clustering has contributed to the regularization term in [116] and [117]. A subspace structure learned from the HSIs was introduced to form a subspace regularizer $\|\mathbf{S} - \mathbf{SH}\|_F^2 + \tau\|\mathbf{H}\|_*$, thereby capturing the global correlation of all pixels [118]. In view of the similarities between the substances, Yuan *et al.* [119] introduced a substance dependence constraint. To be more specific, the substance dependence is considered in the whole space to describe the global spatial information, which is reflected by the abundance. In [120], spectral–spatial joint sparse NMF (S^2-NMF) was proposed by combining the global spatial information and local spectral information simultaneously. Based on the label information, Jia *et al.* [115] developed a similarity regularizer to compensate the dissimilarity regularizer.

IV. Structured NMF

A. Weighted NMF

NMF-based methods achieve the unmixing by utilizing the statistical properties of HSIs. Thus, it is closely related to the number of samples. However, there is a large difference for the number of pixels concerning different endmembers in many scenarios. Therefore, a clusterwise weighted NMF (CW-NMF) [121] method was provided to explore the information of imbalanced pixels. In particular, a weight matrix based on the result of clustering is integrated into the NMF, expressed as

$$\min_{\mathbf{A},\mathbf{S}} \|(\mathbf{X} - \mathbf{AS})\mathbf{W}\|_F^2 \tag{25}$$

where $\mathbf{W}$ is the diagonal weight matrix. Then, the effects of the pixels involving imbalanced endmembers can be enhanced by

giving larger weight values, while the effects of the pixels that involving majority endmembers are reduced by assigning smaller weight values. Similarly, self-paced NMF (SpNMF) [122] was presented by adopting weighted least-squares losses, given as

$$\min_{\mathbf{A},\mathbf{S}} \sum_{b=1}^{B} \left\{ w_b \|\mathbf{x}^b - (\mathbf{AS})^b\|_2^2 + h(\gamma, w_b) \right\} \tag{26}$$

where h is a self-paced function associated with "model age" parameter γ to learn the weights adaptively. There are many self-paced functions, such as binary, linear, logarithmic, and mixture functions. Besides, the SpNMF can be also extended for pixel weighting (i.e., $\min_{\mathbf{A},\mathbf{S}} \sum_{p=1}^{P} \{w_p \|\mathbf{x}_p - (\mathbf{AS})_p\|_2^2 + h(\gamma, w_p)\}$) and element weighting (i.e., $\min_{\mathbf{A},\mathbf{S}} \sum_{b,p=1}^{B,P} \{w_{bp}[x_{bp} - (\mathbf{AS})_{bp}]^2 + h(\gamma, w_{bp})\}$). Meanwhile, the weighted NMF is also effective to improve the robustness of NMF.

B. Projection-Based NMF

Inspired by the sparse regression-based unmixing methods, a projection-based NMF (PNMF) algorithm was proposed by introducing a spectral library into the NMF framework [123]. In detail, related spectra in the library $\mathbf{U}$ are projected onto a subspace based on a transformation matrix $\mathbf{V}$ to obtain the projected endmembers, i.e., $\mathbf{A} = \mathbf{UV}$, the cost function is

$$\min_{\mathbf{V},\mathbf{S}} \|\mathbf{X} - \mathbf{UVS}\|_F^2. \tag{27}$$

In this way, the endmembers are not only adaptively generated from the spectral library, but also related to the HSIs. Meanwhile, the number of endmembers, which is a difficulty in sparse regression, is controlled by the dimension of the subspace.

C. Convex NMF

In NMF, some of the computed endmembers in $\mathbf{A}$ may be artificial, which do not belong to any real material in the scene, but exist only in the solution space of the problem. To keep the association between extracted and real endmembers, each spectrum $\mathbf{a}_m$ was assumed to be nonnegative linear combinations of the observed pixels [124], i.e., $\mathbf{A} = \mathbf{X\Xi}$. Thus, the cost function is

$$\min_{\mathbf{\Xi},\mathbf{S}} \|\mathbf{X} - \mathbf{X\Xi S}\|_F^2. \tag{28}$$

Along with subspace clustering, Lu *et al.* [114] also proposed to extract endmembers by linearly combining of all pixels in a spectral subspace, avoiding the generation of artificial endmembers. Under this assumption, the hierarchical sparse NMF (HSNMF) [58] introduced hierarchical sparsity constraints to accommodate endmember variability.

D. Robust NMF

NMF is applied to the unmixing under the assumption of Gaussian noise. However, HSIs are inevitably contaminated by different types of noise, e.g., Gaussian noise, impulse noise, stripes, and deadlines. Hence, the classical NMF model defined by the least-squares loss is sensitive to noise, leading to dramatically degrading the unmixing performance. To improve the robustness of NMF, many models have been reported based on certain metrics, including but not limited to bounded Itakura–Saito (IS) divergence [125], $L_{2,1}$-norm regularizer [62], [113], [126], [127], CIM [90], [94], [128], [129], Cauchy function [130], and general robust loss function [131]. The bounded IS divergence was employed to address the additive, multiplicative, and mixed noises in HSIs [125].

The $L_{2,1}$-norm regularizer proposed by Ding *et al.* [168] can effectively handle noise and outliers. Based on this regularizer, different robust models are designed to reduce the impact of noise. In [62], robust NMF (rNMF) was proposed by introducing a residual term $\mathbf{E} \in \mathbb{R}^{B\times P}$ accounting for outliers (i.e., nonlinear effects), whose cost function is written as

$$\min_{\mathbf{A},\mathbf{S},\mathbf{E}} D\left(\mathbf{X}|(\mathbf{AS}+\mathbf{E})\right) + \lambda\|\mathbf{E}\|_{2,1} \tag{29}$$

where $D(\mathbf{X}|(\mathbf{AS}+\mathbf{E}))$ measures dissimilarity between $\mathbf{X}$ and $(\mathbf{AS}+\mathbf{E})$, $\|\mathbf{E}\|_{2,1} = \sum_{p=1}^{P} \|\mathbf{e}_p\|_2$, and $\mathbf{e}_p$ denotes the pth vector of $\mathbf{E}$. As such, the problem (29) addresses the nonlinearities and improves the robustness against the noisy pixels. Based on the $L_{1,2}$-norm regularizer, He *et al.* [126] used specific bands of the HSIs and modeled the sparse noise explicitly for significantly improving the robustness of NMF. The $L_{2,1}$-norm was employed to replace Euclidean distance or KL divergence directly in [113], and the spectral–spatial constrained NMF (SS-NMF) model was developed to cope with non-Gaussian noises or outliers, whose reconstruction error is calculated by

$$\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|_{2,1}. \tag{30}$$

In this way, the model is robust to noisy pixels by columns. To further achieve robustness to band noise by rows, the $L_{1,2}$-norm was also incorporated to form a spectral–spatial robust NMF model for hyperspectral unmixing [127].

Correntropy is a nonlinear similarity measure, which is based on Gaussian kernel

$$\kappa(x) = \frac{1}{2\pi\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right). \tag{31}$$

Recently, CIM is employed to replace the least-squares loss to develop some robust models. For example, a correntropy-based NMF (CENMF) [128] was proposed to suppress the influence of noisy bands efficiently. Considering the diversity of the noise levels of pixels, correntropy-based spatial–spectral robust sparsity-regularized NMF (CSsRS-NMF) was proposed in [94] by adaptive assigning weights to noisy pixels. Furthermore, robustness can be achieved from an elementwise noise perspective [90], [129].

By cutting off the large error via the truncation operation, the truncated Cauchy loss [169] exhibits robustness to outliers. Accordingly, the reconstruction error was measured via the truncated Cauchy function [130], which is expressed as

$$h(x) = \begin{cases} \ln\left(1 + \frac{x^2}{\gamma^2}\right), & |x| \le \epsilon \\ \ln\left(1 + \frac{\epsilon^2}{\gamma^2}\right), & |x| \ge \epsilon \end{cases} \tag{32}$$

where γ and ϵ are the scale parameter and truncated coefficient, respectively.

Furthermore, a general loss-based NMF (GLNMF) model [131] was developed by introducing a general robust loss function defined in [170], given as

$$f(x) = \frac{|2-\upsilon|}{\upsilon}\left(\left(\frac{(x/c)^2}{|2-\upsilon|}+1\right)^{\upsilon/2}-1\right) \tag{33}$$

to downplay the large noise. It is a generalization of the L_2 loss ($\upsilon \to 2$), Cauchy loss ($\upsilon \to 0$), and Welsch loss ($\upsilon \to -\infty$), etc.

Through the $L_{2,1}$-norm, the model can effectively address the pixels with noise or outliers. By contrast, CIM can be applied flexibly to relieve the impact of noise from bands, pixels, and elements insights. Meanwhile, the algorithms with CIM are effective to process non-Gaussian and impulsive noise. The truncated Cauchy loss can suppress the outliers effectively.

V. Generalized NMF

A. Linear-Quadratic NMF

Combined with the linear-quadratic model, NMF can be extended as $\mathbf{X} = \mathbf{AS} = \mathbf{A}_a\mathbf{S}_a + \mathbf{A}_b\mathbf{S}_b$. Subsequently, the following cost function requires to be minimized:

$$\min_{\mathbf{A}_a,\mathbf{A}_b,\mathbf{S}_a,\mathbf{S}_b} \|\mathbf{X} - \mathbf{A}_a\mathbf{S}_a - \mathbf{A}_b\mathbf{S}_b\|_F^2. \tag{34}$$

Under this framework, the unmixing task can be addressed from the following two perspectives.

1) First, Tang *et al.* [132] proposed an unmixing method by using the prior knowledge of some signatures in the scene. To be specific, $\mathbf{A}_a$ and $\mathbf{A}_b$ represent the matrices of known and unknown endmembers with related abundance fractions $\mathbf{S}_a$ and $\mathbf{S}_b$, respectively.
2) Then, the second-order scattering of light can be considered to improve the performance [61], [133]–[135]. In this case, $\mathbf{A}_a$ and $\mathbf{A}_b$ are the endmember matrix and bilinear endmember matrix, while $\mathbf{S}_a$ and $\mathbf{S}_b$ are abundance matrix and interaction abundance matrix, respectively. Among them, semi-nonnegative matrix factorization (semi-NMF) was used for the optimization to process an entire image in the matrix form [61], [135].

B. Kernelized NMF

Kernel methods can be introduced for nonlinear hyperspectral unmixing without estimating the nonlinear mixture model. In [136], a constrained kernel NMF (CKNMF) was proposed for dealing with nonlinearities. Through kernel mappings (denoted by ϕ), the observed matrix $\mathbf{X}$ and endmember matrix $\mathbf{A}$ are transformed to high-dimensional feature space, obtaining $\phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_N)]$ and $\phi(\mathbf{A}) = [\phi(\mathbf{a}_1), \ldots, \phi(\mathbf{a}_M)]$. As a result, the data are linearly separable in the feature space. The cost function is given as

$$\min_{\mathbf{A},\mathbf{S}} \|\phi(\mathbf{X}) - \phi(\mathbf{A})\mathbf{S}\|_F^2 \tag{35}$$

where ϕ is a nonlinear function. Generally, the Gaussian kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$ is utilized to achieve dot product operator in a high-dimensional kernel feature space. Taking into account both the input and feature spaces, a biobjective NMF [137] was formulated by combining the linear and kernel-based models. However, the kernel-based methods suffer from computation burden. For dealing with large-scale and streaming dynamic data, Zhu *et al.* [138] proposed an online KNMF (OKNMF) framework to control the computational complexity via adopting the stochastic gradient descent (SGD), mini-batch, and averaged SGD strategies. In addition, the KNMF was extended to incremental KNMF (IKNMF) and improved IKNMF (IIKNMF) for desired unmixing accuracy and efficiency [34].

C. Transform Domain-Based NMF

Due to the dense spectra (typically 10 nm) and overlapped information, HSIs are compressible in a suitable transformed domain. Very recently, wavelet transform has been exploited to express the hyperspectral data compactly [139]. Specifically, biorthogonal discrete wavelet transform was employed to represent the hyperspectral data in the wavelet domain, denoted as $\mathbf{x}_{pw}$. Accordingly, the LMM in the wavelet domain can be written as $\mathbf{X}_\mathrm{w} = \mathbf{A}_\mathrm{w}\mathbf{S}_\mathrm{w} + \mathbf{N}_\mathrm{w}$. Hence, the cost function is

$$\min_{\mathbf{A}_\mathrm{w},\mathbf{S}_\mathrm{w}} \|\mathbf{X}_\mathrm{w} - \mathbf{A}_\mathrm{w}\mathbf{S}_\mathrm{w}\|_F^2. \tag{36}$$

On this basis, three prior terms (i.e., volume regularizer, spatial smoothness prior, and sparseness constraint) in the wavelet domain are integrated to better handle the ill-posedness. Similarly, a wavelet-based approach was proposed for estimating abundances in [171]. Furthermore, considering that the curvelet is capable of characterizing anisotropic singularity, such as curves or edges in the image, Xu *et al.* [93] adopted fast discrete curvelet transform to impose a sparse regularizer on the transformed domain of abundances, thereby enhancing both sparsity and diversity.

D. Nonnegative Tensor Factorization (NTF)

A third-order tensor, which is the high-dimensional extension of a matrix, can be used to represent the HSI for preserving the intrinsic structure information. Accordingly, NTF has been successfully applied to HSIs processing, such as denoising [172] and classification [173]. In [140], the NTF method was used to the unmixing task by using canonical polyadic decomposition. However, a limitation is the lack of an explicit link with LMM. In [141], a matrix-vector NTF (MV-NTF) unmixing method was proposed based on block term decomposition. The MV-NTF factorizes the HSI tensor into a sum of several component tensors as

$$\mathcal{X} = \sum_{m=1}^{M} \mathbf{S}_m \circ \mathbf{a}_m + \mathcal{G} = \sum_{m=1}^{M} \left(\mathbf{C}_m\mathbf{B}_m^T\right) \circ \mathbf{a}_m + \mathcal{G} \tag{37}$$

where $\mathcal{X} \in \mathbb{R}^{r\times c\times B}$ is a third-order HSI tensor, $\mathbf{a}_m$ is regarded as an endmember, $\mathbf{S}_m$ is the corresponding abundances represented by the product of two low-rank matrices $\mathbf{C}_m \in \mathbb{R}^{r\times R}$ and $\mathbf{B}_m^T \in \mathbb{R}^{R\times c}$, and $\mathcal{G}$ denotes the noise term. R is related to the rank of abundance matrix, and $\circ$ denotes the outer product.

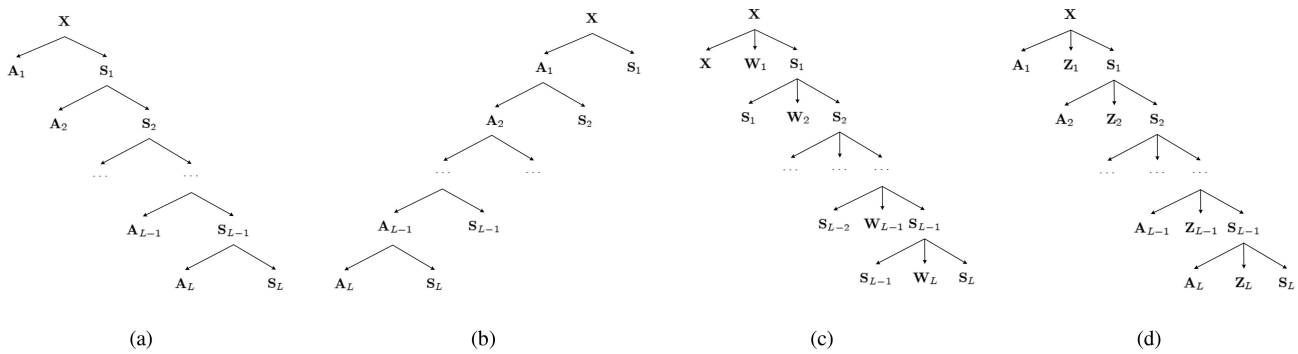


Fig. 3. Architecture of decomposition for (a) multilayer NMF by factorizing the coefficient matrix, (b) multilayer NMF by factorizing the basis matrix, (c) multilayer convex NMF, and (d) multilayer nsNMF.

Apparently, this model constructs a straightforward link between LMM and tensor factorization. The cost function is expressed as

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\mathcal{X} - \sum_{m=1}^{M} \left(\mathbf{C}_m \mathbf{B}_m^T\right) \circ \mathbf{a}_m\|_F^2. \tag{38}$$

Although intrinsic structure information is preserved, the local spatial information is not fully exploited due to the strict rank constraint. Subsequently, under this framework, Xiong *et al.* [143] further incorporated the similarity graph regularizer [142] and the TV regularizer so as to describe the local spatial information. Likewise, three constraints were embedded into MV-NTF, including sparsity, minimum volume, and nonlinearity in [144]. In addition, some unmixing methods were also presented by combining additional constraints [145]–[147]. Considering that NMF characterizes more local spatial details through dealing with HSI at the pixel level, MV-NTF and NMF were coupled to make full use of their individual advantages [148]. In [149], a low-rank tensor regularization was introduced during the learning process, allowing flexibility to the rank of the estimated abundance tensor. Endmember variability was considered based on the 4-D endmember tensor that was constrained by a new low-rank regularization [59], [150]. Besides, a nonlocal Tucker decomposition method [151] was provided to exploit the spectral–spatial correlations and the nonlocal self-similarity.

E. Multilayer/Deep Extensions

The aforementioned approaches explore the information in a single-layer manner, which do not allow for hierarchical refinement of the obtained endmembers and abundances. In order to extract hierarchical features with hidden information, DL [39] has achieved commendable success in pattern recognition [174]–[176]. Hence, Rajabi and Ghassemian [52] unfolded NMF into multilayer architecture (i.e., multiple basis matrices and one abundance matrix) and proposed the multilayer NMF (MLNMF) model for hyperspectral unmixing. As shown in Fig. 3(a), in the first layer, the matrix $\mathbf{X}$ can be factorized into $\mathbf{A}_1$ and $\mathbf{S}_1$. Then, in the next layer, $\mathbf{S}_1$ is further factorized into $\mathbf{A}_2$ and $\mathbf{S}_2$. A similar process is continued until the factorization of the Lth layer is completed. Here, L represents the maximum number of layers. Accordingly, the observation matrix is decomposed into $L+1$ nonnegative factors, i.e., $\mathbf{X} = \mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_L\mathbf{S}_L$. The latent factors can be obtained by minimizing the following cost function:

$$\min_{\mathbf{A}_l\mathbf{S}_l} \|\mathbf{S}_{l-1} - \mathbf{A}_l\mathbf{S}_l\|_F^2 \tag{39}$$

with $l = 1, \cdots, L$, and $\mathbf{S}_0 = \mathbf{X}$ when $l = 1$. Then, the endmember and abundance matrices are $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_L$ and $\mathbf{S} = \mathbf{S}_L$, respectively. Based on this multilayer architecture, a double-constrained multilayer NMF (DCMLNMF) [152] method was proposed to jointly explore the sparsity and the geometric structure. Besides, Tong *et al.* [153] developed novel unmixing approaches by further imposing constraints, such as a spectral and spatial total variation regularizer, an adaptive graph regularizer [154], and a homogeneous region regularizer [155]. Moreover, the classical MLNMF was restructured and improved by integrating the Hoyer's projector in [35]. Different from factorizing $\mathbf{S}$ mentioned previously, $\mathbf{A}$ is decomposed in [156] to form constrained multilayer NMF (CMLNMF) along with MVC and sparsity constraints, as shown in Fig. 3(b). Thus, $\mathbf{X} = \mathbf{A}_L\mathbf{S}_L\cdots\mathbf{S}_2\mathbf{S}_1$. Here, the endmember and abundance matrices are $\mathbf{A} = \mathbf{A}_L$ and $\mathbf{S} = \mathbf{S}_L\cdots\mathbf{S}_2\mathbf{S}_1$, respectively. Furthermore, multilayer factorization was investigated with fast kernel archetypal analysis (AA) [157] and kernel NMF [158] for unmixing.

However, these models are optimized by only minimizing the cost function of each layer, which fail to reduce the total reconstruction error. Trigeorgis *et al.* [176] formulated deep Semi-NMF, achieving a significant breakthrough. Motivated by this, a deep NMF structure [53] was proposed to address the unmixing task, whose cost function is

$$\min_{\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_L,\mathbf{S}_L} \|\mathbf{X} - \mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_L\mathbf{S}_L\|_F^2 \tag{40}$$

where $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_L$ and $\mathbf{S} = \mathbf{S}_L$ denote the endmembers and abundances, respectively. The proposed model consists of *pretraining stage* and *fine-tuning stage*, where the former pretrains all factors layer by layer and the latter is used to reduce the total reconstruction error. Likewise, a sparsity-constrained deep NMF (L_1-DNMF) was proposed for hyperspectral unmixing [159]. Multiview concept learning was incorporated to

explicitly model the consistent and complementary information [177].

In addition, an asymmetric *encoder–decoder* framework was presented in [160] for hyperspectral unmixing, where a multilayer nonlinear network was designed to powerfully encode the original data, and the resulting abundances were then decoded by the decoder part with one layer. The cost function is given by

$$\min_{\mathbf{A},\mathbf{E},\mathbf{W}_1,\ldots,\mathbf{W}_L} \|\mathbf{X}-\mathbf{E}-\mathbf{A}\sigma(\mathbf{W}_L\cdots\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{X})))\|_F^2 \tag{41}$$

where $\sigma(\cdot)$ is a nonlinear activation function, abundances $\mathbf{S}=\mathbf{S}_L=\sigma(\mathbf{W}_L\cdots\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{X})))$, and $\mathbf{E}$ is introduced in the *decoder* to characterize sparse noise.

Besides, various deep NMF models [178] have been proposed in an increasing number of applications.

1) Deep orthogonal NMF [178]–[180] is a variant of deep NMF by imposing orthogonality constraint to $\mathbf{S}_l$, where the decomposition is the same as in [156].
2) Deep convex NMF [175] was developed by extending convex NMF [181] that is also known as AA or concept factorization (CF), where each basis vector (named concept) is modeled as a linear combination of data points, i.e., $\mathbf{a}_m=\sum_{n=1}^{N}\mathbf{x}_n w_{nm}$. Fig. 3(c) shows the factorization process of multilayer CF [182]. Accordingly, $\mathbf{X}=\mathbf{X}\mathbf{W}_1\mathbf{S}_1\cdots\mathbf{W}_L\mathbf{S}_L$. After layerwise factorization, the cost function of deep convex NMF is

$$\min_{\mathbf{W}_1,\ldots,\mathbf{W}_L,\mathbf{S}_1,\ldots,\mathbf{S}_L} \|\mathbf{X}-\mathbf{X}\mathbf{W}_1\mathbf{S}_1\cdots\mathbf{W}_L\mathbf{S}_L\|_F^2. \tag{42}$$

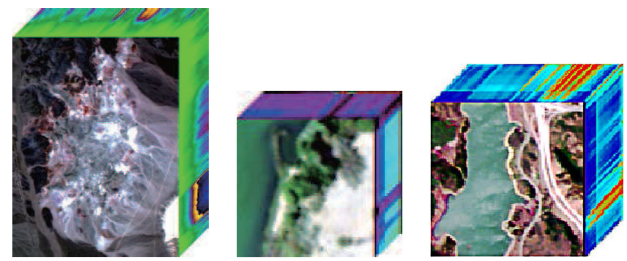
Different from directly using the output of the previous layer as the input of subsequent layer, Zhang *et al.* [183] proposed a novel deep self-representative concept factorization network (DSCF-Net) and a deep semisupervised coupled factorization network (DS^{2}CF-Net) [184].
3) By introducing a smoothing matrix at each layer, deep nsNMF (dnsNMF) [185] was reported to learn features hierarchically in the context of text mining. Let $\mathbf{Z}_l$ denote the "smoothing" matrix. The matrix $\mathbf{X}$ is factorized into $\mathbf{A}_1$, $\mathbf{Z}_1$, and $\mathbf{S}_1$ in the first layer. Then, $\mathbf{S}_1$ is factorized into $\mathbf{A}_2$, $\mathbf{Z}_2$, and $\mathbf{S}_2$ in the next layer. The same process will be continued until the Lth layer is reached, as shown in Fig. 3(d). Hence, the observation can be represented using $\mathbf{X}=\mathbf{A}_1\mathbf{Z}_1\cdots\mathbf{A}_L\mathbf{Z}_L\mathbf{S}_L$. The cost function of dnsNMF is expressed as

$$\min_{\mathbf{A}_1,\mathbf{Z}_1,\ldots,\mathbf{A}_L,\mathbf{Z}_L,\mathbf{S}_L} \|\mathbf{X}-\mathbf{A}_1\mathbf{Z}_1\cdots\mathbf{A}_L\mathbf{Z}_L\mathbf{S}_L\|_F^2. \tag{43}$$

4) Deep autoencoder-like NMF (DANMF) [186] was proposed for community detection, whose cost function is expressed as

$$\min_{\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_L,\mathbf{S}_L} \|\mathbf{X}-\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_L\mathbf{S}_L\|_F^2 + \|\mathbf{S}_L-\mathbf{A}_L^T\cdots\mathbf{A}_2^T\mathbf{A}_1^T\mathbf{X}\|_F^2. \tag{44}$$

Similar to deep autoencoder, DANMF consists of an *encoder* component and a *decoder* component. This architecture empowers DANMF to learn the hierarchical mappings between the original network and the final community assignment with implicit low-to-high level hidden attributes of the original network learned in the intermediate layers.



(a) (b) (c)

Fig. 4. Visualization of (a) Cuprite image, (b) Samson image, and (c) Jasper Ridge image.

Meanwhile, the hierarchical factorization has applied to tensor, developing some multilayer frameworks of tensor decomposition [187], and deep tensor decompositions [188]–[190]. In addition, the deep unfolding technique was used for unrolling the iteration inference algorithm into a layerwise structure to obtain novel neural network-like architectures that enjoy the advantages of well-defined interpretability, strong learning power, and little computational cost [51], [161], [162].

Multilayer/deep extensions of NMF combine both interpretability and the extraction of multiple hierarchical features. Nevertheless, it is also an important and challenging research issue such as how to determine the parameters (e.g., the inner ranks and the number of layers) and loss function, and how to choose efficient optimization algorithms and initial conditions.

VI. Experimental Results and Discussion

SAD is utilized to assess unmixing performance quantitatively, given as

$$\mathrm{SAD}_m=\arccos\left(\frac{\mathbf{A}_m^T\hat{\mathbf{A}}_m}{\|\mathbf{A}_m^T\|\|\hat{\mathbf{A}}_m\|}\right) \tag{45}$$

where $\mathbf{A}_m$ and $\hat{\mathbf{A}}_m$ are the mth original and estimated endmember spectral signatures, respectively.

We choose the most popular methods in different categories to conduct experiments, such as $L_{1/2}$-NMF [84], SGSNMF [97],[1] TV-RSNMF [89], $L_{1/2}$-RNMF [126], MV-NTF-TV [143], ML-NMF [52],[2] and SSRDMF [160]. It should be noted that all the results are averaged after ten independent runs. All experiments are conducted under the environment of MATLAB R2015b software and computer configuration Intel Core i5 CPU at 2.80 GHz and 8.00 GB RAM. For hyperspectral unmixing, the number of endmembers is a crucial factor, which can be set manually or estimated through an effective method, such as virtual dimensionality [191], [192] and HySime [193].

Three hyperspectral scene have been utilized in the tests, shown in Fig. 4. The first dataset is the notable Cuprite data,

[1] [Online]. Available: https://github.com/Xinyu-Wang/SGSNMF_TGRS
[2] [Online]. Available: https://github.com/roozbehrajabi/mlnmf

TABLE II
SAD (AVERAGE OF TEN RUNS) ALONG WITH STANDARD DEVIATION ON THE AVIRIS CUPRITE DATASET FOR DIFFERENT METHODS

Methods	$L_{1/2}$-NMF	SGSNMF	TV-RSNMF	$L_{1/2}$-RNMF	MV-NTF-TV	MLNMF	SSRDMF
Alunite GDS82 Na82	0.1340±6.29%	0.1219±**1.72%**	0.1012±3.12%	**0.0891**±2.04%	0.1047±2.76%	0.1132±3.42%	0.1165±2.72%
Andradite WS487	**0.0731**±1.66%	0.0945±5.49%	0.0756±1.42%	0.0832±2.14%	0.0894±2.44%	0.0748±**1.40%**	0.0936±1.84%
Buddingtonite GDS85 D-206	0.1032±3.20%	**0.0903**±**1.60%**	0.0976±2.37%	0.1009±2.48%	0.0941±2.27%	0.0935±2.55%	0.1281±2.19%
Chalcedony CU91-6A	0.1504±**1.53%**	0.1495±1.64%	0.1420±2.34%	0.1500±1.59%	0.1505±1.67%	0.1539±1.57%	**0.1269**±1.86%
Kaolin/Smect H89-FR-5 30K	0.1004±3.83%	**0.0631**±**1.20%**	0.0689±1.53%	0.0699±1.39%	0.0845±1.83%	0.0914±3.31%	0.0789±1.96%
Kaolin/Smect KLF508 85%K	0.1163±3.78%	0.0903±**1.47%**	0.1099±4.25%	0.1091±3.34%	0.1057±2.98%	0.1190±3.81%	**0.0901**±2.72%
Kaolinite KGa-2	0.1520±5.67%	0.1348±4.16%	0.1621±5.19%	0.1300±3.62%	**0.1283**±4.12%	0.1547±5.91%	0.1486±**2.90%**
Montmorillonite+Illi CM37	0.0583±1.19%	0.0569±2.41%	0.0564±**0.95%**	**0.0528**±1.07%	0.0662±2.60%	0.0563±1.18%	0.0617±1.06%
Muscovite IL107	0.1045±2.59%	0.1047±1.44%	0.0975±3.27%	0.1212±4.11%	**0.0870**±1.82%	0.0924±**0.93%**	0.1025±2.10%
Nontronite NG-1.a	0.1261±1.34%	0.1186±1.18%	0.1263±1.73%	0.1292±**1.10%**	0.1403±4.48%	0.1273±1.37%	**0.1033**±1.47%
Pyrope WS474	0.1134±2.55%	0.1220±**0.99%**	0.1222±3.35%	0.0993±4.47%	0.0876±2.97%	0.0810±2.34%	**0.0721**±1.67%
Sphene HS189.3B	0.0680±1.69%	0.0742±**0.65%**	0.0681±1.03%	0.0829±2.00%	**0.0606**±0.83%	0.0716±1.80%	0.0668±0.84%
Mean	0.1083±0.81%	0.1017±0.60%	0.1023±0.90%	0.1015±0.61%	0.0999±**0.41%**	0.1024±0.57%	**0.0991**±0.53%

Boldfaced number denotes the best result under each condition.

TABLE III
SAD SCORES (AVERAGE OF TEN RUNS) ALONG WITH THEIR STANDARD DEVIATION ON THE SAMSON DATASET FOR DIFFERENT METHODS

Methods	$L_{1/2}$-NMF	SGSNMF	TV-RSNMF	$L_{1/2}$-RNMF	MV-NTF-TV	MLNMF	SSRDMF
Soil	0.0620±11.28%	**0.0085**±**0.00%**	0.0275±0.47%	0.0281±0.39%	0.0228±0.58%	0.0754±16.39%	0.0320±0.94%
Tree	0.0647±5.03 %	0.0402±**0.01%**	0.0481±0.41%	0.0514±0.30%	0.0474±0.55%	0.0590±3.87 %	**0.0376**±0.32%
Water	0.1167±2.24 %	0.1364±**0.02%**	0.1086±0.14%	0.0998±0.67%	0.1114±0.18%	0.1001±0.66 %	**0.0903**±1.52%
Mean	0.0811±4.72 %	0.0617±**0.01%**	0.0614±0.34%	0.0598±0.45%	0.0605±0.22%	0.0781±6.91 %	**0.0533**±0.46%

Boldfaced number denotes the best result under each condition.

TABLE IV
SAD SCORES (AVERAGE OF TEN RUNS) ALONG WITH THEIR STANDARD DEVIATION ON THE JASPER RIDGE DATASET FOR DIFFERENT METHODS

Methods	$L_{1/2}$-NMF	SGSNMF	TV-RSNMF	$L_{1/2}$-RNMF	MV-NTF-TV	MLNMF	SSRDMF
Tree	0.1959±9.56 %	**0.0818**±**0.63 %**	0.1854±26.21%	0.1349±7.84 %	0.2091±3.98 %	0.1456±8.52 %	0.1518±4.44 %
Water	0.1548±6.56 %	0.1783±6.31 %	**0.0991**±2.78 %	0.1069±3.01 %	0.3443±**1.90** %	0.1421±5.19 %	0.2061±7.75 %
Soil	0.2661±25.31%	0.3738±19.67%	0.1780±16.56%	0.2090±29.59%	0.2497±21.54%	0.1576±21.42%	**0.1223**±**3.18 %**
Road	0.5453±8.19 %	0.3112±31.86%	0.4957±**3.36 %**	0.4870±8.70 %	**0.2474**±9.52%	0.6736±11.68%	0.4355±12.45%
Mean	0.2905±7.36 %	0.2363±**3.54 %**	0.2395±8.09 %	0.2345±9.30 %	0.2626±4.91 %	0.2797±4.09 %	**0.2289**±4.03 %

Boldfaced number denotes the best result under each condition.

which was obtained by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). The second dataset is Samson image, which is the first public real hyperspectral dataset. The third dataset is Jasper Ridge scene. Since these real HSI datasets have been studied for unmixing in many articles, we refer to the experimental results reported in some published references.

1) *AVIRIS Cuprite Dataset:* As shown in Fig. 4(a), it is typically used to verify the effectiveness of hyperspectral unmixing methods. It contains 250×191 pixels, and each pixel contains 188 bands (i.e., 3–103, 114–147, and 168–220) after removal of noisy bands. The covered wavelength range comprises $0.4 - 2.5\mu$m. There are mainly 12 minerals in the subscene: Alunite GDS82 Na82, Andradite WS487, Buddingtonite GDS85 D-206, Chalcedony CU91-6A, Kaolin/Smect H89-FR-5 30K, Kaolin/Smect KLF508 85%K, Kaolinite KGa-2, Montmorillonite + Illi CM37, Muscovite IL107, Nontronite NG-1.a, Pyrope WS474, and Sphene HS189.3B. The reference signatures are from the U.S. Geological Survey (USGS) spectral library,[3] and a mineral map is often used for illustrative purposes.[4]

2) *Samson Dataset:* As shown in Fig. 4(b), it contains 95×95 pixels, and each pixel has 156 bands ranging from 0.401 to 0.889μm. The number of endmembers is set to 3, including soil, tree, and water.

3) *Jasper Ridge Dataset:* As shown in Fig. 4(c), it contains 100×100 pixels, and each pixel has 198 bands (i.e., 4–107, 113–153, and 167–219) after removal of noisy bands ranging from 0.38 to 2.5μm. The number of endmembers is set to 4, including tree, water, soil, and road.

The experimental results are summarized in Tables II–IV and plotted in Figs. 5–10. Tables II, III, and IV list the SAD values between each reference spectrum and the endmembers

[3][Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06
[4][Online]. Available: https://www.usgs.gov/labs/spectroscopy-lab

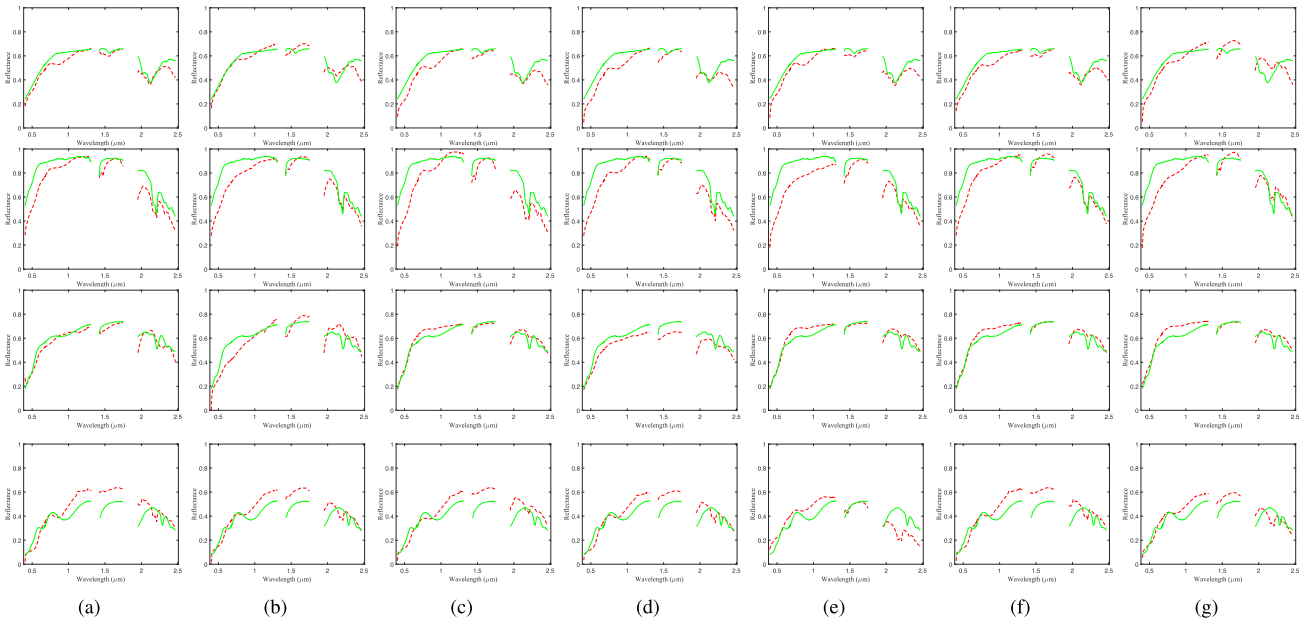


Fig. 5. Comparison of the USGS library spectra (green solid line) with those estimated by different methods (red dash line) of four endmembers on the AVIRIS Cuprite dataset. From top to bottom: Buddingtonite GDS85 D-206, Kaolinite KGa-2, Montmorillonite+Illi CM37, and Nontronite NG-1.a. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

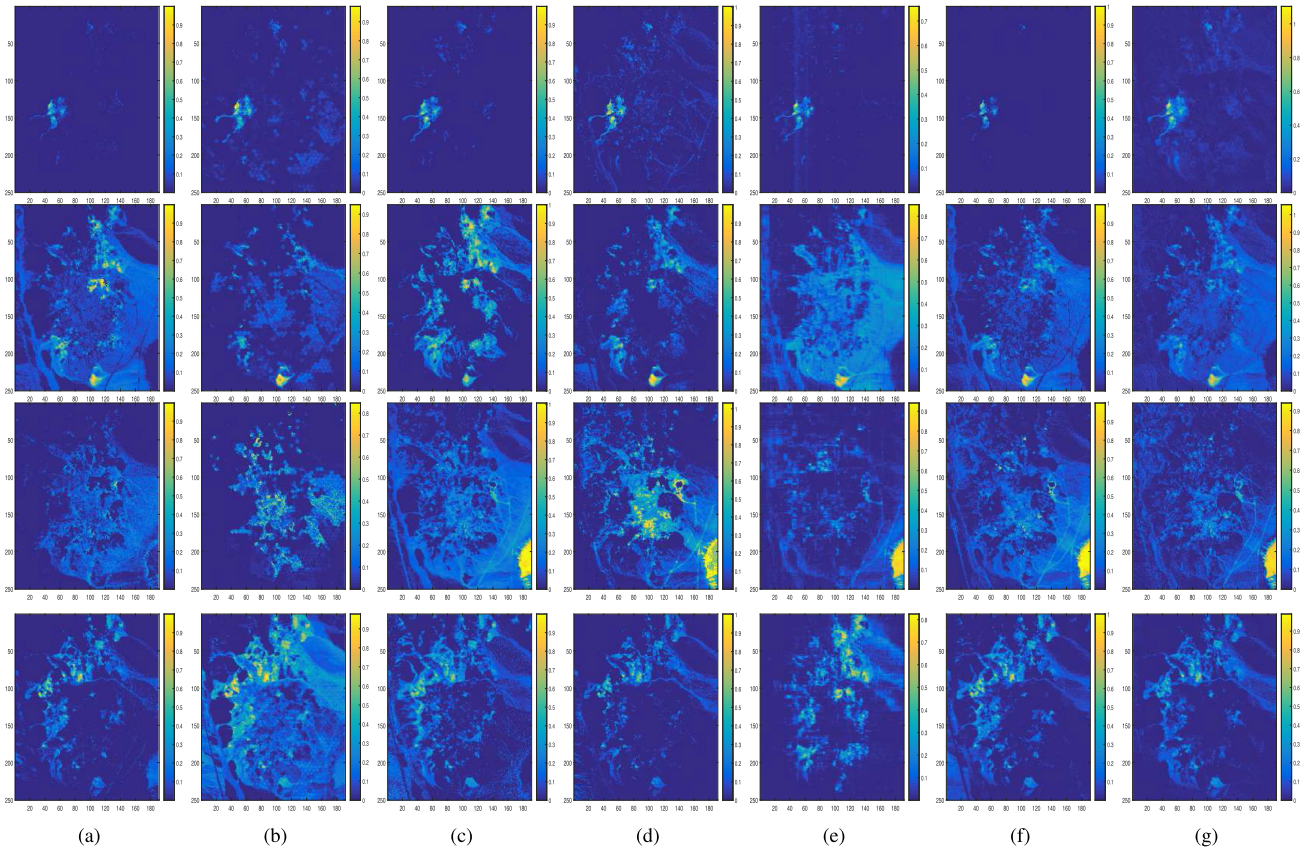


Fig. 6. Fractional abundance maps estimated by different methods of four endmembers on the AVIRIS Cuprite dataset. From top to bottom: Buddingtonite GDS85 D-206, Kaolinite KGa-2, Montmorillonite+Illi CM37, and Nontronite NG-1.a. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

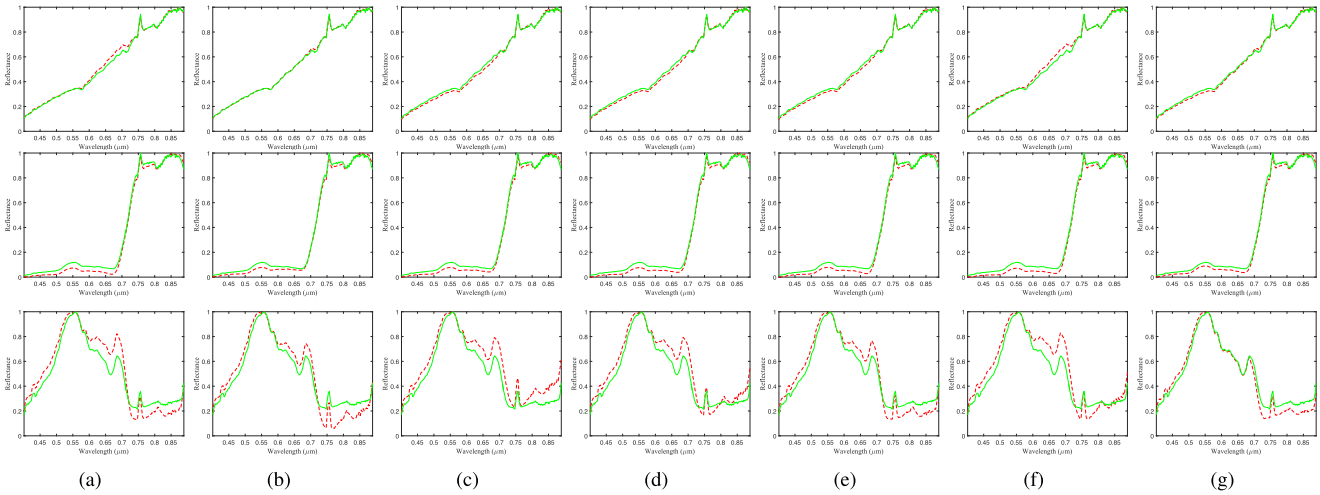


Fig. 7. Comparison of the reference spectra (green solid line) with those estimated by different methods (red dash line) of three endmembers on the Samson dataset. From top to bottom: soil, tree, and water. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

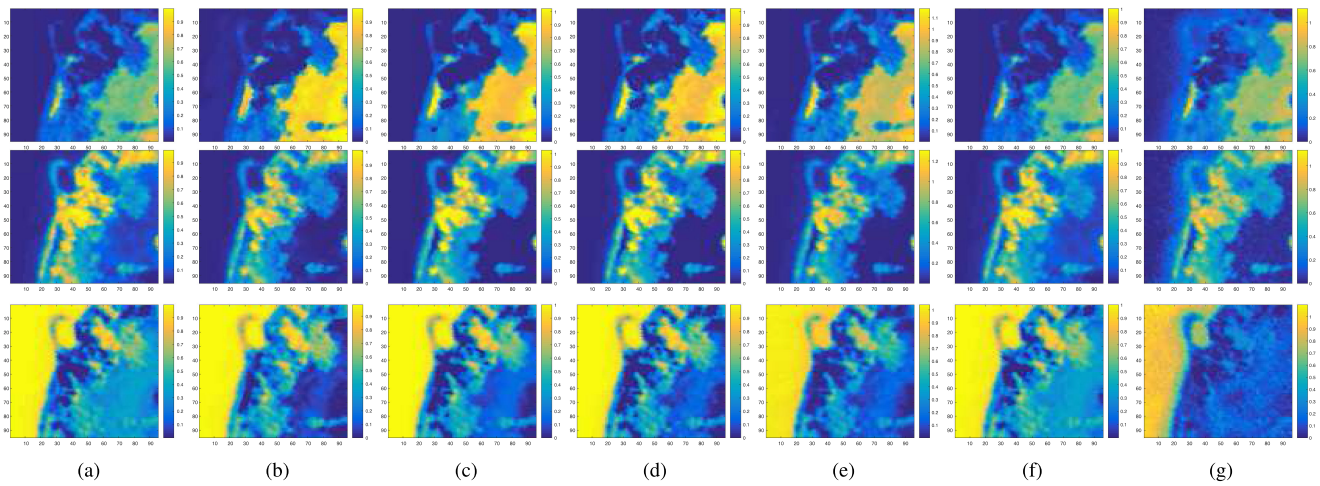


Fig. 8. Fractional abundance maps estimated by different methods of three endmembers on the Samson dataset. From top to bottom: soil, tree, and water. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

extracted by each method on Cuprite, Samson, and Jasper Ridge datasets, respectively. Figs. 5, 7, and 9 present the correlation of the endmember signatures obtained by these seven methods and the reference signatures. Figs. 6, 8, and 10 show the visual comparison of abundance maps estimated by all algorithms.

From Tables II–IV, all methods achieve satisfactory SAD mean values for most materials. It demonstrates that the introductions of sparsity regularizer, spatial information, robust constraint, and multilayer/deep architectures in hyperspectral unmixing are compelling. Besides, compared with $L_{1/2}$-NMF, methods SGSNMF, TV-RSNMF, and MV-NTF-TV achieve lower mean SAD scores, revealing the significance of preserving the spatial information. $L_{1/2}$-RNMF aims to model the sparse noise explicitly, thereby obtaining desirable unmixing results. The performance of MLNMF and SSRDMF verifies that multilayer/deep architectures offer significant advantage. In addition, the results obtained by SSRDMF are better than those provided by MLNMF. This indicates that the combination of multilayer nonlinear network and self-supervised constraint can play a significant role in the task of improving unmixing performance. In particular, from the standard deviation perspective, SGSNMF often provides the best results since the endmember matrix is initialized based on the results of the segmentation and the region-based VCA.

For illustrative purposes, Figs. 5, 7, and 9 plot the reference signatures (green solid line) along with the endmembers identified by different methods (red dash line) on the Cuprite, Samson, and Jasper Ridge datasets, respectively. It can be seen that the endmember signatures are nearly all around correlated in spectral terms with respect to the reference partners. Meanwhile, Figs. 6, 8, and 10 show the abundance maps estimated by different methods on the Cuprite, Samson, and Jasper

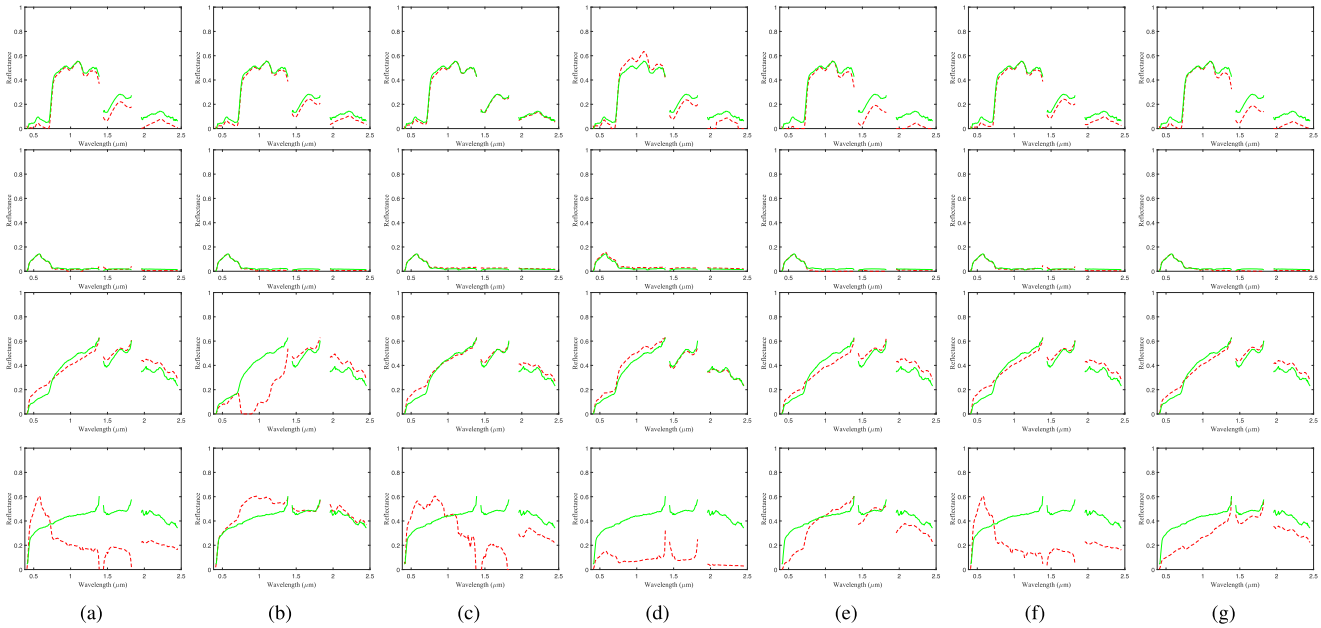


Fig. 9. Comparison of the reference spectra (green solid line) with those estimated by different methods (red dash line) of three endmembers on the Jasper Ridge dataset. From top to bottom: tree, water, soil, and road. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

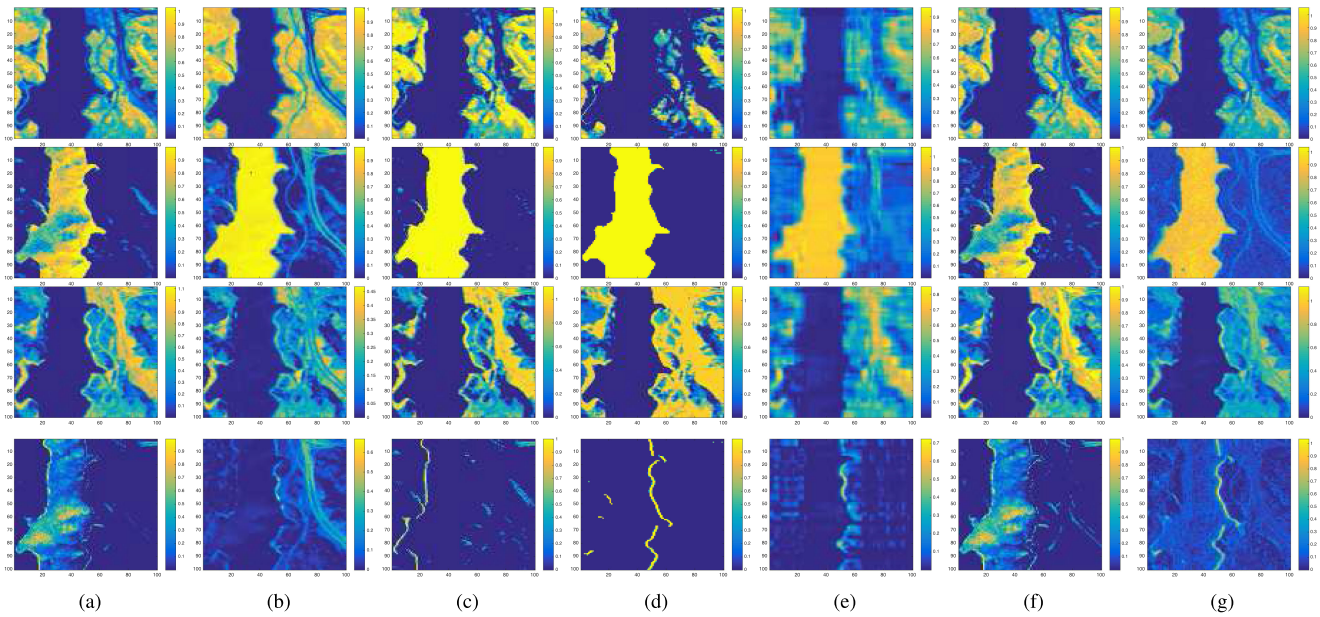


Fig. 10. Fractional abundance maps estimated by different methods of three endmembers on the Jasper Ridge dataset. From top to bottom: tree, water, soil, and road. From left to right: (a) $L_{1/2}$-NMF, (b) SGSNMF, (c) TV-RSNMF, (d) $L_{1/2}$-RNMF, (e) MV-NTF-TV, (f) MLNMF, and (g) SSRDMF.

Ridge datasets, respectively. Due to the ground-truth abundance maps are unavailable, we compare the obtained abundance maps with the real scene image, i.e., Fig. 4. In the abundance maps, the higher value means the proportion of the material is larger. From Fig. 6, it is obvious that similar abundance maps can be achieved in most cases for each material. From Fig. 8, we can observe all estimation results have a good correlation with the geological maps of the Samson dataset. For the Jasper Ridge dataset, as shown in Figs. 9 and 10, it is difficult to obtain desirable estimations in terms of Road for all methods, which may be because there are few pixels for Road so that VCA fails to extract this signature when initializing.

These seven methods have their own benefits. Figs. 6(a), 8(a), and 10(a) show that $L_{1/2}$-NMF can obtain a sparse abundance map because $L_{1/2}$-norm regularizer is an optimal choice for

the hyperspectral unmixing. Figs. 6(b), 8(b), and 10(b) keep the spatial group structure and the sparsity within a local spatial group through utilizing the spatial group sparsity regularizer. The results in Figs. 6(c), 8(c), and 10(c) show that it is effective to utilize the piecewise smoothness. The estimated abundances in Figs. 6(d), 8(d), and 10(d) are robust to noise, which is mainly because $L_{1/2}$-RNMF describes the sparse noise explicitly. Figs. 6(e), 8(e), and 10(e) preserve the local spatial structure and the global spectral–spatial information. The results listed in Figs. 6(f)–(g), 8(f)–(g), and 10(f)–(g) display that the abundance maps estimated by MLNMF and SSRDMF are always in accordance with the ground truth, demonstrating the effectiveness of the multilayer/deep architectures. From the results of experiments, we find that sparse regularizer, spectral–spatial information, multilayer/deep architectures are all beneficial to the hyperspectral unmixing.

Last, we investigate the average running time of ten times on Jasper Ridge dataset for $L_{1/2}$-NMF, SGSNMF, TV-RSNMF, $L_{1/2}$-RNMF, MV-NTF-TV, MLNMF, and SSRDMF, which are 4.99, 60.81, 25.19, 43.31, 107.50, 19.38, and 136.54 s, respectively. Apparently, the $L_{1/2}$-NMF performs the fastest estimation since it is a single-layer factorization with efficiency sparse regularizer. SGSNMF and TV-RSNMF require more time due to the learning of spatial group structure and piecewise smooth structure, respectively. To handle sparse noise, $L_{1/2}$-RNMF models the sparse noise explicitly, increasing the running time. MLNMF is also fast since it only decomposes the observation matrix iteratively layer by layer. MV-NTF-TV and SSRDMF need much more time mainly owing to the complex factorization.

VII. Discussion and Future Directions

NMF plays an increasingly significant role in the field of hyperspectral unmixing. In particular, the constrained NMF has the capacity of providing more accurate endmembers and abundances by integrating the spectral constraints and the spatial constraints. The structured NMF enables flexibility to account for more structures and details, such as the difference of pixels, bands, and elements. By extending the decomposition form, the generalized NMF exhibits great potential in acquiring more essential characteristics, e.g., nonlinearities, 3-D structure information, and hidden information.

Nevertheless, there are still several drawbacks. For instance, the constrained NMF generally requires extensive parameter tunning to achieve satisfactory results. The generalized NMF often suffers from time-consuming. Secondly, the NMF-based methods rely on proper guidance or initialization to generate meaningful endmembers. Besides, it is difficult to thoroughly capture the plentiful information of HSIs. To this end, it is quite challenging to obtain excellent unmixing performance. In the future, how to design NMF methods for unmixing deserves further research. We provide some considerations as follows.

1) One important research direction is to make use of the spatial and spectral information simultaneously to guarantee a more reliable unmixing performance. The spectral–spatial joint has shown considerable potentials for hyperspectral unmixing.
2) Many NMF approaches have been reported to exploit the spectral characteristics, such as the corresponding simplex volume, endmember distance, and signature smoothness. However, more efforts are required to describe the endmember variability under the NMF model, e.g., constructing a 4-D endmember tensor in [59]. Combining insight from spectral variability with a mathematical treatment would be valuable to improve the performance significantly.
3) Most of the current NMF algorithms rely strongly on LMM to obtain unmixing results. In real scenarios, multipath scattering is common due to complex landforms, resulting in nonlinear spectral mixture effects. As shown in [61] and [133]–[135], a nonlinear mixture model is closely related to the LMM, indicating NMF methods that aim to solving nonlinear unmixing problems deserve to be further investigated.
4) To achieve more reliable performance in practical scenarios, there is growing attention on improving the robustness of the methods. Although many robust NMF methods have been proposed, they mainly focus on the robustness to noise. In addition to relieve the effect of different types of noise, it is also important to investigate the robustness to the selections of the initialization methods and the tunable parameters in various application scenarios.
5) The computational complexity is also an aspect that brings difficulties to apply most existing methods (e.g., graph regularized algorithms, NTF, and multilayer/deep approaches). Meanwhile, HSIs are very large in general. Considering that many applications need real- or near real-time processing, it is crucial to develop fast alternatives to reduce the computation time.

References

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[2] G. Swayze *et al.*, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. JPL Airborne Earth Sci. Workshop*, 1992, pp. 47–49.
[3] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Aug. 1998.
[4] 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.
[5] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
[6] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "Abundance estimation using discontinuity preserving and sparsity-induced priors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2148–2158, Jul. 2019.
[7] J. Senthilnath, S. N. Omkar, V. Mani, N. Karnwal, and S. P. B, "Crop stage classification of hyperspectral data using unsupervised techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 861–866, Apr. 2013.
[8] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.
[9] R. Pu, "Mapping tree species using advanced remote sensing technologies: A state-of-the-art review and perspective," *J. Remote Sens.*, vol. 2021, 2021, Art. no. 26.

[10] A. Gowen, C. O'Donnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging - An emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, Dec. 2007.

[11] Y. Feng and D. Sun, "Application of hyperspectral imaging in food safety inspection and control: A review," *Crit. Rev. Food Sci.*, vol. 52, no. 11, pp. 1039–1058, Nov. 2012.

[12] A. R. Huete, T. Miura, and G. Xiang, "Land cover conversion and degradation analyses through coupled soil-plant biophysical parameters derived from hyperspectral EO-1 hyperion," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1268–1276, Aug. 2003.

[13] C. Gendrin, Y. Roggo, and C. Collet, "Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review," *J. Pharmaceut. Biomed.*, vol. 48, no. 3, pp. 533–553, Nov. 2008.

[14] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[15] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Ann. JPL Airborne Geosci. Workshop*, 1993, pp. 11–14.

[16] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Imag. Spectrom. V*, 1999, pp. 266–277.

[17] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *Proc. 21st Can. Symp. Remote Sens.*, 1999, pp. 21–24.

[18] J. M. P. Nascimento and J. M. Bioucas-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.

[19] C. Chang, C. Wu, W. Liu, and Y. Ouyang, "A growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.

[20] C. Chang, H. Li, C. Wu, and M. Song, "Recursive geometric simplex growing analysis for finding endmembers in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 296–308, Jan. 2017.

[21] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[22] J. M. Bioucas-Dias, "A variable splitting augmented lagragian approach to linear spectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2009, pp. 1–4.

[23] B. Yang, Z. Chen, and B. Wang, "Nonlinear endmember identification for hyperspectral imagery via hyperpath-based simplex growing and fuzzy assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 351–366, Jan. 2020.

[24] C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1651–1679, Aug. 2017.

[25] M.-D. 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.

[26] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[27] R. Wang, H.-C. Li, A. Pizurica, J. Li, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using double reweighted sparse regression and total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1146–1150, Jul. 2017.

[28] S. Zhang, J. Li, H.-C. Li, C. Deng, and A. Plaza, "Spectral-spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3265–3276, Jun. 2018.

[29] 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.

[30] L. Qi, J. Li, Y. Wang, Y. Huang, and X. Gao, "Spectral-spatial-weighted multiview collaborative sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8766–8779, Dec. 2020.

[31] T. Ince, "Superpixel based graph laplacian regularization for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 5501305.

[32] N. Wang, B. Du, L. Zhang, and L. Zhang, "An abundance characteristic-based independent component analysis for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 416–428, Jan. 2015.

[33] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[34] R. Huang, X. Li, and L. Zhao, "Hyperspectral unmixing based on incremental kernel nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6645–6662, Nov. 2018.

[35] Y. Yuan, Z. Zhang, and G. Liu, "A novel hyperspectral unmixing model based on multilayer NMF with Hoyer's projection," *Neurocomputing*, vol. 440, pp. 145–158, Jan. 2021.

[36] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Sep. 2009.

[37] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[38] J. S. Bhatt, M. V. Joshi, and M. S. Raval, "A data-driven stochastic approach for unmixing hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1936–1946, Jun. 2014.

[39] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computat.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[40] J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2189–2192.

[41] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1427–1431, Sep. 2018.

[42] B. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Hyperspectral unmixing using a neural network autoencoder," *IEEE Access*, vol. 6, pp. 25646–25656, May 2018.

[43] 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.

[44] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[45] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1467–1471, Sep. 2019.

[46] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "Spectral unmixing using autoencoder with spatial and spectral regularizations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3321–3324.

[47] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Blind hyperspectral unmixing using autoencoders: A critical comparison," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1340–1372, 2022.

[48] D. Hong *et al.*, "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 28, 2021, doi: 10.1109/TNNLS.2021.3082289.

[49] M. Zhao, L. Yan, and J. Chen, "LSTM-DNN based autoencoder network for nonlinear hyperspectral image unmixing," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 2, pp. 295–309, Feb. 2021.

[50] V. S. S., V. S. Deshpande, and J. S. Bhatt, "A practical approach for hyperspectral unmixing using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 5511505.

[51] C. Zhou and M. R. D. Rodrigues, "ADMM-based hyperspectral unmixing networks for abundance and endmember estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022, Art. no. 5520018.

[52] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 38–42, Jan. 2015.

[53] X.-R. Feng, H.-C. 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.

[54] B. Somers, S. Delalieux, J. Stuckens, W. W. Verstraeten, and P. Coppin, "A weighted linear spectral mixture analysis approach to address endmember variability in agricultural production systems," *Int. J. Remote Sens.*, vol. 30, no. 1, pp. 139–147, Jan. 2009.

[55] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Enviro.*, vol. 115, no. 7, pp. 1603–1616, Jul. 2011.

[56] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[57] R. A. Borsoi *et al.*, "Spectral variability in hyperspectral data unmixing: A comprehensive review," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 223–270, Dec. 2021.

[58] T. Uezato, M. Fauvel, and N. Dobigeon, "Hierarchical sparse nonnegative matrix factorization for hyperspectral unmixing with spectral variability," *Remote Sens.*, vol. 12, pp. 1–24, 2020.

[59] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "Low-rank tensor modeling for hyperspectral unmixing accounting for spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1833–1842, Mar. 2020.

[60] S. E. Brezini, M. S. Karoui, F. Z. Benhalouche, Y. Deville, and A. Ouamri, "An NMF-based method for hyperspectral unmixing using a structured additively-tuned linear mixing model to address spectral variability," in *Proc. Mediterranean Middle-East Geosci. Remote Sens. Symp.*, 2020, pp. 45–48.

[61] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.

[62] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[63] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. 13th Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[64] J. Cao, L. Zhuo, and H. Tao, "An endmember initialization scheme for nonnegative matrix factorization and its application in hyperspectral unmixing," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 5, May 2018, Art. no. 195.

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

[66] S. A. Robila and L. G. Maciak, "Considerations on parallelizing nonnegative matrix factorization for hyperspectral data unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 57–61, Jan. 2009.

[67] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: An optimal gradient method for nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.

[68] 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.

[69] 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, 2020.

[70] 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.

[71] 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.

[72] 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, 2020.

[73] Y. Yu, S. Guo, and W. Sun, "Minimum distance constrained nonnegative matrix factorization for the endmember extraction of hyperspectral images," in *Proc. SPIE MIPPR 2007: Remote Sens. GIS Data Process. Appl., Innov. Multispectral Technol. Appl.*, 2007, pp. 151–159.

[74] Z. Wu, S. Ye, J. Liu, L. Sun, and Z. Wei, "Sparse non-negative matrix factorization on GPUs for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3640–3649, Aug. 2014.

[75] B. Yang, W. Luo, and B. Wang, "Constrained nonnegative matrix factorization based on particle swarm optimization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3693–3710, Aug. 2017.

[76] B. Yang, B. Wang, and Z. Wu, "Unsupervised nonlinear hyperspectral unmixing based on bilinear mixture models via geometric projection and constrained nonnegative matrix factorization," *Remote Sens.*, vol. 10, no. 5, pp. 1–34, May 2018.

[77] C. Revel, Y. Deville, V. Achard, X. Briottet, and C. Weber, "Inertia-constrained pixel-by-pixel nonnegative matrix factorisation: A hyperspectral unmixing method dealing with intra-class variability," *Remote Sens.*, vol. 10, no. 11, Nov. 2018, Art. no. 1706.

[78] 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.

[79] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[80] 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.

[81] 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.

[82] 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. 6531–6544, Nov. 2016.

[83] W. Wang, Y. Qian, and H. Liu, "Multiple clustering guided nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5162–5179, 2020.

[84] 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.

[85] 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.

[86] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, Dec. 2014.

[87] R. Huang, X. Li, and L. Zhao, "Nonnegative matrix factorization with data-guided constraints for hyperspectral unmixing," *Remote Sens.*, vol. 9, no. 10, Oct. 2017, Art. no. 1074.

[88] 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.

[89] 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.

[90] T. Peng, Z. Wu, Z. Chen, and C. Deng, "Robust nonnegative local coordinate factorization for hyperspectral unmixing," in *Proc. IEEE 3rd Int. Conf. Signal Image Process.*, 2018, pp. 306–309.

[91] C. G. Tsinos, A. A. Rontogiannis, and K. Berberidis, "Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 160–174, Jun. 2017.

[92] X. Xu, J. Li, S. Li, and A. Plaza, "Generalized morphological component analysis for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2817–2832, Apr. 2020.

[93] X. Xu, J. Li, S. Li, and A. Plaza, "Curvelet transform domain-based sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4908–4924, 2020.

[94] X. Li, R. Huang, and L. Zhao, "Correntropy-based spatial-spectral robust sparsity-regularized hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1453–1471, Feb. 2021.

[95] F. Xiong, J. Zhou, J. Lu, and Y. Qian, "Nonconvex nonseparable sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6088–6100, 2020.

[96] S. Khoshsokhan, R. Rajabi, and H. Zayyani, "Sparsity-constrained distributed unmixing of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1279–1288, Apr. 2019.

[97] 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.

[98] 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.

[99] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 757–772, Feb. 2011.

[100] J. Liu, J. Zhang, Y. Gao, C. Zhang, and Z. Li, "Enhancing spectral unmixing by local neighborhood weights," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1545–1552, Oct. 2012.

[101] R. Liu, B. Du, and L. Zhang, "Hyperspectral unmixing via double abundance characteristics constraints based NMF," *Remote Sens.*, vol. 8, no. 6, pp. 1–23, Jun. 2016.

[102] 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.

[103] 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.

[104] B. Rathnayake, E. M. M. B. Ekanayake, K. Weerakoon, G. M. R. I. Godaliyadda, M. P. B. Ekanayake, and H. M. V. R. Herath, "Graph-based blind hyperspectral unmixing via nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6391–6409, Sep. 2020.

[105] T. Peng and W. Tan, "A manifold Hessian-regularized NMF for hyperspectral data unmixing," *Remote Sens. Lett.*, vol. 11, no. 1, pp. 86–95, Jan. 2020.

[106] 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.

[107] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 101–118, Feb. 2014.

[108] Z. Zhang, S. Liao, H. Zhang, S. Wang, and Y. Wang, "Bilateral filter regularized l_2 sparse nonnegative matrix factorization for hyperspectral unmixing," *Remote Sens.*, vol. 10, no. 6, Jun. 2018, Art. no. 816.

[109] M. Wang, B. Zhang, X. Pan, and S. Yang, "Group low-rank nonnegative matrix factorization with semantic regularizer for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1022–1029, Apr. 2018.

[110] W. Wang, Y. Qian, and 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.

[111] 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.

[112] J. Zhang, X. Zhang, X. Tang, P. Chen, and L. Jiao, "Sketch-based region adaptive sparse unmixing applied to hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8840–8856, Dec. 2020.

[113] G. Zhang, S. Mei, Y. Feng, and Q. Du, "Spectral-spatial constrained nonnegative matrix factorization for spectral mixture analysis of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6766–6776, 2021.

[114] X. Lu, L. Dong, and Y. Yuan, "Subspace clustering constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3007–3019, May 2020.

[115] Y. Jia, S. Kwong, J. Hou, and W. Wu, "Semi-supervised non-negative matrix factorization with dissimilarity and similarity regularization," *IEEE Trans. Neural Netw. Learn. Sys.*, vol. 31, no. 7, pp. 2510–2521, Jul. 2020.

[116] 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.

[117] B. Du, S. Wang, N. Wang, L. Zhang, D. Tao, and L. Zhang, "Hyperspectral signal unmixing based on constrained non-negative matrix factorization approach," *Neurocomputing*, vol. 204, pp. 153–161, Sep. 2016.

[118] L. Zhou *et al.*, "Subspace structure regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4257–4270, 2020.

[119] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.

[120] L. Dong, Y. Yuan, and X. Lu, "Spectral-spatial joint sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2391–2402, Mar. 2021.

[121] X. Lv, W. Wang, and H. Liu, "Cluster-wise weighted NMF for hyperspectral images unmixing with imbalanced data," *Remote Sens.*, vol. 13, no. 2, Jan. 2021, Art. no. 268.

[122] 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.

[123] Y. Yuan, Y. Feng, and X. Lu, "Projection-based NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2636–2643, Jun. 2015.

[124] N. Akhtar and A. Mian, "RCMF: Robust constrained matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3354–3366, Jun. 2017.

[125] C. Li, X. Chen, and Y. Jiang, "On diverse noises in hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5388–5402, Oct. 2015.

[126] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.

[127] 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.

[128] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4027–4040, Nov. 2015.

[129] 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.*, vol. 19, pp. 1–5, 2022, Art. no. 5500505.

[130] H. Wang, W. Yang, and N. Guan, "Cauchy sparse NMF with manifold regularization: A robust method for hyperspectral unmixing," *Knowl.-Based Syst.*, vol. 184, Nov. 2019, Art. no. 16.

[131] J. Peng, W. Sun, F. Jiang, H. Chen, Y. Zhou, and Q. Du, "A general loss-based nonnegative matrix factorization for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 5500105.

[132] W. Tang, Z. Shi, and Z. An, "Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures," *Opt. Eng.*, vol. 51, no. 8, pp. 087001–08700110, Aug. 2012.

[133] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, and X. Briottet, "Linear-quadratic blind source separation using NMF to unmix urban hyperspectral images," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1822–1833, Apr. 2014.

[134] O. Eches and M. Guillaume, "A bilinear-bilinear nonnegative matrix factorization method for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 778–782, Apr. 2014.

[135] X. Zhang, J. Zhang, C. Li, C. Cheng, L. Jiao, and H. Zhou, "Hybrid unmixing based on adaptive region segmentation for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3861–3875, Jul. 2018.

[136] X. Li, J. Cui, and L. Zhao, "Blind nonlinear hyperspectral unmixing based on constrained kernel nonnegative matrix factorization," *Signal, Image Video Process.*, vol. 8, no. 8, pp. 1555–1567, Nov. 2014.

[137] F. Zhu and P. Honeine, "Biobjective nonnegative matrix factorization: Linear versus kernel-based models," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4012–4022, Jul. 2016.

[138] F. Zhu and P. Honeine, "Online kernel nonnegative matrix factorization," *Signal Process.*, vol. 131, pp. 143–153, Feb. 2017.

[139] V. S. S. and J. S. Bhatt, "A blind spectral unmixing in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10287–10302, 2021.

[140] M. A. Veganzones, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, "Nonnegative tensor CP decomposition of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2577–2588, May 2016.

[141] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017.

[142] F. Xiong, J. Chen, J. Zhou, and Y. Qian, "Superpixel-based nonnegative tensor factorization for hyperspectral unmixing," in *Proc. 38th IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 6392–6395.

[143] 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.

[144] B. Feng and J. Wang, "Constrained nonnegative tensor factorization for spectral unmixing of hyperspectral images: A case study of urban impervious surface extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 583–587, Apr. 2019.

[145] H.-C. Li, S. Liu, X.-R. Feng, R. Wang, and Y.-J. Sun, "Double weighted sparse nonnegative tensor factorization for hyperspectral unmixing," *Int. J. Remote Sens.*, vol. 42, no. 8, pp. 3180–3191, Apr. 2021.

[146] J.-J. Wang *et al.*, "Endmember independence constrained hyperspectral unmixing via nonnegative tensor factorization," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106657.

[147] P. Zheng, H. Su, and Q. Du, "Sparse and low-rank constrained tensor factorization for hyperspectral image unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1754–1767, 2021.

[148] 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, 2020.

[149] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "A low-rank tensor regularization strategy for hyperspectral unmixing," in *Proc. IEEE Statist. Signal Process. Workshop*, 2018, pp. 373–377.

[150] L. Sun and H. Guo, "Blind unmixing of hyperspectral images based on L_1 norm and tucker tensor decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 5508605.

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

[152] 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, Mar. 2019.

[153] L. Tong *et al.*, "Spectral and spatial total-variation-regularized multilayer non-negative matrix factorization for hyperspectral unmixing," *J. Appl. Remote Sens.*, vol. 13, no. 3, Sep. 2019, Art. no. 1.

[154] 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, 2020.

[155] L. Tong, B. Qian, J. Yu, and C. Xiao, "Homogeneous region regularized multilayer non-negative matrix factorization for hyperspectral unmixing," *J. Appl. Remote Sens.*, vol. 14, no. 4, Dec. 2020, Art. no. 046502.

[156] L. Chen, S. Chen, and X. Guo, "Multilayer NMF for blind unmixing of hyperspectral imagery with additional constraints," *Photogrammetric Eng. Remote Sens.*, vol. 83, no. 4, pp. 307–316, Apr. 2017.

[157] G. Zhao, C. Zhao, and X. Jia, "Multilayer unmixing for hyperspectral imagery with fast kernel archetypal analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1532–1536, Oct. 2016.

[158] J. Liu, Y. Zhang, Y. Liu, and C. Mu, "Hyperspectral images unmixing based on abundance constrained multi-layer KNMF," *IEEE Access*, vol. 9, pp. 91080–91090, 2021.

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

[160] H.-C. Li, X.-R. Feng, D.-H. Zhai, Q. Du, and A. Plaza, "Self-supervised robust deep matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 5513214.

[161] Y. Qian, F. Xiong, Q. Qian, and J. Zhou, "Spectral mixture model inspired network architectures for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7418–7434, Oct. 2020.

[162] F. Xiong, J. Zhou, S. Tao, J. Lu, and Y. Qian, "SNMF-Net: Learning a deep alternating neural network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5510816.

[163] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 403–415, Mar. 2006.

[164] Y. Liu, Y. Guo, F. Li, L. Xin, and P. Huang, "Sparse dictionary learning for blind hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 578–582, Apr. 2019.

[165] J. Huang, T.-Z. Huang, L.-J. Deng, and X.-L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2437, Apr. 2019.

[166] L. Drumetz, T. R. Meyer, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3435–3450, Jul. 2019.

[167] D. Cai, X. He, J. Han, and T. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[168] C. Ding, D. Zhou, X. He, and H. Zha, "R1-PCA:Rotational invariant l_1-norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 281–288.

[169] N. Guan, T. Liu, Y. Zhang, D. Tao, and L. S. Davis, "Truncated cauchy nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 246–259, Jan. 2019.

[170] J. T. Barron, "A general and adaptive robust loss function," 2017. [Online]. Available: http://arxiv.org/abs/1701.03077

[171] V. S. S. and J. S. Bhatt, "A novel approach for abundance estimation in wavelet domain," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2159–2162.

[172] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogram. Remote Sens.*, vol. 83, pp. 50–63, Sep. 2013.

[173] Z. Zhong *et al.*, "Discriminant tensor spectral-spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1028–1032, May 2015.

[174] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[175] B. Qian, X. Shen, Z. Tang, and T. Zhang, "Deep convex NMF for image clustering," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 583–590.

[176] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[177] W. Zhao, C. Xu, Z. Guan, and Y. Liu, "Multiview concept learning via deep matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 814–825, Feb. 2021.

[178] P. D. Handschutter, N. Gillis, and X. Siebert, "A survey on deep matrix factorizations," *Comput. Sci. Rev.*, vol. 58, pp. 1–18, 2021, Art. no. 100423.

[179] B. Lyu, K. Xie, and W. Sun, "A deep orthogonal non-negative matrix factorization method for learning attribute representations," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 443–452.

[180] Y. Qiu, G. Zhou, and K. Xie, "Deep approximately orthogonal nonnegative matrix factorization for clustering," 2017, *arXiv:1711.07437*.

[181] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

[182] X. Li, X. Shen, Z. Shu, Q. Ye, and C. Zhao, "Graph regularized multilayer concept factorization for data representation," *Neurocomputing*, vol. 238, pp. 139–151, May 2017.

[183] Y. Zhang *et al.*, "Deep self-representative concept factorization network for representation learning," in *Proc. SIAM Int. Conf. Data Mining*, 2020, pp. 361–369.

[184] Y. Zhang *et al.*, "Dual-constrained deep semi-supervised coupled factorization network with enriched prior," *Int. J. Comput. Vis.*, vol. 129, no. 12, pp. 3233–3254, Dec. 2021.

[185] J. Yu, G. Zhou, A. Cichocki, and S. Xie, "Learning the hierarchical parts of objects by deep non-smooth nonnegative matrix factorization," *IEEE Access*, vol. 6, pp. 58096–58105, 2018.

[186] F. Ye, C. Chen, and Z. Zheng, "Deep autoencoder-like nonnegative matrix factorization for community detection," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1393–1402.

[187] X. Bi, A. Qu, and X. Shen, "Multilayer tensor factorization with applications to recommender systems," *Ann. Statist.*, vol. 46, no. 6, pp. 3308–3333, 2018.

[188] C. Jia, M. Shao, and Y. Fu, "Sparse canonical temporal alignment with deep tensor decomposition for action recognition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 738–750, Feb. 2017.

[189] S. Oymak and M. Soltanolkotabi, "End-to-end learning of a convolutional neural network via deep tensor decomposition," 2018, *arXiv:1805.06523*.

[190] J. Casebeer, M. Colomb, and P. Smaragdis, "Deep tensor factorization for spatially-aware scene decomposition," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 180–184.

[191] C.-I. Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, Apr. 2018.

[192] V. S. S., J. S. Bhatt, and B. Chattopadhyay, "Virtual dimensionality of hyperspectral data: Use of multiple hypothesis testing for controlling Type-I error," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2974–2985, 2020.

[193] 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.

Xin-Ru Feng received the B.Sc. degree in communication engineering in 2016 from the Southwest Jiaotong University, Chengdu, China, where she is currently working toward the Ph.D. degree in signal and information processing with the School of Information Science and Technology.

Her research focuses on hyperspectral image analysis and processing.

Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

He is currently a Full Professor with the School of Information Science and Technology, Southwest Jiaotong University. From November 2013 to October 2014, he has been a Visiting Scholar working with Prof. William J. Emery with the University of Colorado Boulder, Boulder, CO, USA. He has authored more than 140 research papers, more than 90 JCR journal papers, such as published in the ISPRS *Journal of Photogrammetry and Remote Sensing*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Pattern Recognition*. His research interests include statistical analysis of SAR images, remote sensing image processing, and pattern recognition.

Prof. Li was the recipient of the 2021 Sichuan Natural Science Award, 2021 CSIG Natural Science Award, and 2018 Best Reviewer Award from IEEE Geoscience and Remote Sensing Society for his service to IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is an Associate Editor for the IEEE JSTARS, and is an Editorial Board Member of *Journal of Southwest Jiaotong University* and *Journal of Radars*. He was a Guest Editor of Special Issues of the *Journal of Real-Time Image Processing,* IEEE JSTARS, and IEEE JOINT MODELING AND SIMULATION SYSTEM, a Program Committee Member for the 26th International Joint Conference on Artificial Intelligence (IJCAI-2107) and the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTITEMP-2019), and the Session Chair for the 2017 International Geoscience and Remote Sensing Symposium (IGARSS-2017) and the 2019 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR-2019). He is also an Active Reviewer of more than 30 international journals.

Rui Wang received the Ph.D. degree in information and communication engineering from Southwest Jiaotong University, Chengdu, China, in 2017.

She is currently an Associate Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University. From September 2015 to September 2016, she was a Visiting Scholar with Ghent University, Ghent, Belgium. Her research interests include remote sensing and image processing.

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

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, 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 the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair with the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was a General Chair for the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She was an Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering and the Graduate Certificate in Higher Education from The University of New South Wales, Sydney, NSW, Australia, in 1996 and 2005, respectively, via part-time study.

She has a lifelong academic career in higher education, which she is always passionate about. She is currently an Associate Professor with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Campbell, ACT, Australia. She has authored or coauthored more than 280 referred papers, including more than 170 journal papers, addressing various topics, including data correction, feature reduction and image classification using machine learning techniques. She has coauthored the remote sensing textbook titled *Remote Sensing Digital Image Analysis*, 3rd ed. (Springer-Verlag, 1999) and 4th ed. (Springer, 2006). These publications are highly cited in the remote sensing and image processing communities. Her research interests include remote sensing, hyperspectral image processing, and spatial data analysis.

Dr. Jia is a Subject Editor of *Journal of Soils and Sediments* and an Associate Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Câceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was the recipient of the Best Column Award of *IEEE Signal Processing Magazine* in 2015 and the Most Highly Cited Paper from 2005 to 2010 in *Journal of Parallel and Distributed Computing*. He was the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President of the Spanish Chapter of the IEEE GRSS from 2012 to 2016. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2013 to 2017 and guest-edited ten special issues on hyperspectral remote sensing for different journals. He is also an Associate Editor for IEEE ACCESS (also an Outstanding Associate Editor in 2017).