Sparse Unmixing in the Presence of Mixed Noise Using ℓ_0-Norm Constraint and Log-Cosh Loss

Yiu Yu Chan, Xiao-Peng Li, *Member, IEEE*, Jiajie Mai, Chi-Sing Leung, *Senior Member, IEEE*, and Hing Cheung So, *Fellow, IEEE*

Abstract—Over the past two decades, sparse unmixing (SU) has gained significant attention in the realm of hyperspectral imaging. The aims of SU are to seek a subset of spectral signatures and estimate their fractional abundances to represent each mixed spectral pixel. Conventional SU methods often employ the Frobenius norm and thus cannot work satisfactorily in the presence of non-Gaussian noise. Second, the ideal ℓ_0-norm is usually substituted with its convex or nonconvex approximation in most existing algorithms, which may degrade the recovery performance. To address these issues, this article proposes a novel approach, termed sparse unmixing using ℓ_0-norm constraint and log-cosh loss (SUNNING). We exploit the log-cosh function to minimize the fitting errors subject to three constraints, namely, nonnegativity, sum-to-one, and upper bounded ℓ_0-norm. Then, we adopt the projected gradient descent (PGD) framework to solve such an optimization problem. SUNNING includes two alternating steps, gradient descent and nonconvex projection, where an optimality of the solution is guaranteed. Also, we prove the convergence of SUNNING, including the objective value and variable sequence. In addition, to attain higher unmixing accuracy, we exploit the spectral library pruning (SLP) strategy to eliminate inactive endmembers, yielding an improved SUNNING. Experimental results on synthetic and real-world datasets exhibit improved robustness and effectiveness of the suggested methods over the state-of-the-art algorithms. MATLAB code is available at: https://github.com/freeLix-YY/IEEE_TGRS2024_SparseUnmixing_SUNNING_demo

Index Terms—Hyperspectral image (HSI) analysis, ℓ_0-norm, projected gradient descent (PGD), robust optimization, sparse unmixing (SU).

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are recorded by imaging spectrometers. This collected spatial and spectral information are widely used in hyperspectral data analysis, such as HSI classification [1], [2], HSI denoising [3], [4], hyperspectral anomaly detection [5], [6], HSI super-resolution [7], [8], and hyperspectral unmixing [9], [10]. Since the received spectral data encapsulate the spectral mixtures of different surface materials, hyperspectral unmixing plays a prominent role in decomposing the mixed pixels into a set of endmembers and their respective abundance maps.

The linear mixture model (LMM) is a common approach for hyperspectral unmixing [11]. It assumes that each pixel could be expressed as a linear combination of a series of endmembers, with coefficients corresponding to the abundances of each endmember. In practice, two conventional constraints are imposed on the abundance vector, viz., the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC) [12], [13]. In the remote sensing community, numerous unmixing algorithms have been proposed, and most of them can be classified into unsupervised and semi-supervised methods. Unsupervised approaches, such as minimum volume simplex analysis (MVSA) [14], [15] and nonnegative matrix factorization (NMF) [16], [17], perform endmember extraction and abundance estimation simultaneously without a priori knowledge of endmembers and/or existence of pure pixels. Semi-supervised learning focuses on fractional abundance estimation, in which the input of spectral library for various surface materials is required [18], [19].

As hyperspectral data generally lie on a low-rank subspace, the signal spectrum can be linearly modeled, and a few endmembers from hundreds of spectral signatures could approximate the signal [20]. Due to the sparsity property and spatial redundancy of HSI, hyperspectral unmixing is considered as a special form of sparse regression [21]. Inspired by sparse learning, various sparse unmixing (SU) techniques have been proposed and applied to HSI analysis. Specifically, they leverage the prior knowledge of a spectral library to obtain an optimal subset of spectral signatures and their fractional abundances. Representative algorithms include SU via variable splitting augmented Lagrangian and total variation (SUnSAL–TV) [22], spectral–spatial weighted SU (S²WSU) [23], and SU using spectral a priori information (SUnSPI) [24]. All these methods adopt the Frobenius norm (F-norm) to measure the reconstruction error. The F-norm loss function performs well for low-noise HSIs in the absence of outliers, but it cannot guarantee promising unmixing results when both Gaussian and non-Gaussian noises appear.

Diving deeper into a real scenario, HSIs might be susceptible to various kinds of noise [25], [26], [27], including

Gaussian noise, impulsive noise, shot noise, and line strips. Gaussian noise may arise during the image acquisition process due to poor lighting conditions or sensor noise. Line strips appear in HSIs captured by either whisk-broom or push-broom sensors. Shot noise comes from defective pixels in HSIs. These noise types can adversely affect the accuracy of the unmixing results. We group the combination of different noise types as "mixed noise." Since the F-norm does not consider the impact of mixed noise on abundance estimates, the resultant abundance maps deviate significantly from the ground truth. On the other hand, to obtain plausible unmixing effects, the existing SU algorithms require tediously tuning procedures to the regularization parameters.

To tackle the aforementioned issues, in this article, we propose a novel method, termed sparse unmixing using ℓ_0-norm constraint and log-cosh loss (SUNNING). To robustly handle non-Gaussian noise, we adopt the log-cosh function as the data fidelity term. In addition, the ℓ_0-norm constraint is added to our formulation. In doing so, the suggested model has two major advantages: 1) it is able to control the number of endmembers explicitly and 2) it could achieve robust unmixing performance. The SUNNING consists of two alternating operations, gradient descent and nonconvex projection. For the nonconvex projection, the constraint set is comprised of ANC, ASC, and upper bounded ℓ_0-norm. Based on the Karush–Kuhn–Tucker (KKT) conditions, an efficient solver for the nonconvex projection is developed. Moreover, we provide a theoretical convergence analysis of SUNNING. A spectral library pruning (SLP) strategy is also employed to boost up the unmixing performance. In post-processing, we attempt to rule out irrelevant signatures and then re-estimate fractional abundances based on the trimmed library. Our main contributions are summarized as follows.

1) SUNNING handles the constrained SU problem without the regularization term and thus explicit sparsity control is allowed. Moreover, SUNNING–SLP improves the performance with the use of the SLP strategy.
2) The log-cosh function is applied to the objective function, and hence, the proposed method is robust to mixed noise. That is, it obtains satisfactory performance in different noise scenarios.
3) An efficient solver is developed for the nonconvex projection, which guarantees an optimal solution. In addition, we prove that the objective value and variable sequence are convergent.
4) Compared with the existing SU algorithms, the SUNNING achieves better performance on synthetic and real hyperspectral datasets under different kinds of mixed noise.

The rest of this article is organized in the following manner. Section II reviews the relevant literature and various SU formulations. In Section III, we develop the SUNNING algorithm and provide theoretical analysis. We introduce the SLP strategy and propose an improved version of SUNNING in Section IV. Section V presents and discusses numerical simulation results. Ultimately, concluding remarks are given in Section VI.

II. OVERVIEW OF RELATED WORKS

This section provides an overview of a variety of SU formulations. First, HSI with L-bands is represented as a 3-D data cube. Typically, the image in each band is vectorized, and thus, it is reshaped as a matrix. Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_P] \in \mathbb{R}^{L \times P}$ be an observed HSI with L-bands and P pixels, while $\boldsymbol{A} \in \mathbb{R}^{L \times n}$ represents a spectral library consisting of n spectral signatures. In addition, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_P] \in \mathbb{R}^{n \times P}$ is denoted as a fractional abundance matrix. The LMM assumes that the spectrum of each pixel can be expressed as a linear combination of endmembers, and their corresponding abundances exist in each pixel. Therefore, the observed $\boldsymbol{y}_j$ can be modeled as follows:

$$\boldsymbol{y}_j = \boldsymbol{A}\boldsymbol{x}_j + \boldsymbol{e}_j \quad \text{for} \quad j \in [1, P] \tag{1}$$

where $\boldsymbol{e}_j \in \mathbb{R}^L$ is the noise vector. In most cases, the abundance vector $\boldsymbol{x}_j$ is sparse because only a few spectral signatures of $\boldsymbol{A}$ are likely contributing to each pixel. In light of this, the following ℓ_0-norm sparse regression problem is formulated:

$$\min_{\boldsymbol{x}_j} \|\boldsymbol{A}\boldsymbol{x}_j - \boldsymbol{y}_j\|_2^2 \quad \text{s.t.} \ \boldsymbol{x}_j \geq 0, \quad \mathbf{1}^\top \boldsymbol{x}_j = 1, \quad \|\boldsymbol{x}_j\|_0 \leq S \tag{2}$$

where $S > 0$ controls the maximum number of nonzero elements in $\boldsymbol{x}_j$. Both SD-MMV [28] and NNOMP-PGD [29] employ the greedy pursuit algorithm to address (2).

Another popular direction for solving (2) is to apply the penalty function concept, which substitutes the ℓ_0-norm sparsity constraint with a regularizer. In fact, $\boldsymbol{X}$ has the sparsity characteristic among the row vectors, indicating that the number of endmembers contained in the mixed pixels is smaller than spectral samples in the dictionary $\boldsymbol{A}$. To utilize this prior, CLSUnSAL [30] proposes using $\ell_{2,1}$-norm to characterize the row-sparsity property. To extend the idea of collaborative SU, CSUnL0 [31] adopts $\ell_{2,0}$-norm to obtain a sparser solution

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F^2 + \lambda \|\boldsymbol{X}\|_{2,0} + \mathcal{I}_{R_+}(\boldsymbol{X}) \tag{3}$$

where $\lambda > 0$ is the regularization parameter, $\|\boldsymbol{X}\|_{2,0} = \sum_{i=1}^n \mathbf{1}(\|\boldsymbol{x}^{i,:}\|_2 > 0)$ is the $\ell_{2,0}$-norm, $\boldsymbol{x}^{i,:}$ is the ith row vector of $\boldsymbol{X}$, and $\mathcal{I}_{R_+}(\cdot)$ is an indicator function. When $\boldsymbol{X}$ is nonnegative, $\mathcal{I}_{R_+}(\cdot) = 0$. Otherwise, $\mathcal{I}_{R_+}(\cdot) = +\infty$. In addition, surrogate functions to supplant the ℓ_0-norm have been extensively researched. The ℓ_1-norm is popularly adopted as the primary substitute, such as SUnSAL [32] and SUnSAL-TV [22]. To reinforce the sparsity of the ℓ_1-norm, DRSU-TV [33] and DRSGHU [34] propose a double reweighted $\ell_{1,1}$-norm regularizer that simultaneously promotes column sparsity and elementwise sparsity of the abundance $\boldsymbol{X}$. As the columns of abundances have a linear correlation, the low-rank representation has been devised to account for the spatial correlation [35], leading to

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F^2 + \lambda \|\boldsymbol{X}\|_{2,0} + \tau \operatorname{rank}(\boldsymbol{X}) + \mathcal{I}_{R_+}(\boldsymbol{X}) \tag{4}$$

where $\tau > 0$ is another regularization parameter.

Due to the F-norm, the above formulations are susceptible to outliers and/or mixed noise. There are two research pathways to tackle mixed noise, viz., introducing a sparse noise term and employing a robust loss function.

For the former ones, the joint sparsity and total variation-based unmixing (JSTV) method [26] is proposed

$$\min_{\boldsymbol{X},\boldsymbol{E}} \|\boldsymbol{Y} - \boldsymbol{AX} - \boldsymbol{E}\|_F^2 + \lambda_{\text{TV}}\text{TV}(\boldsymbol{X}) + \lambda_2\|\boldsymbol{X}\|_{2,1} + \lambda_3\|\boldsymbol{E}\|_1 \tag{5}$$

where $\text{TV}(\cdot)$ is the 2-D total variation operator that represents total variation along horizontal and vertical directions applied to $\boldsymbol{X}$, $\|\boldsymbol{X}\|_{2,1} = \sum_{i=1}^n \|\boldsymbol{x}^{i,:}\|_2$ is the $\ell_{2,1}$-norm, and $\|\boldsymbol{E}\|_1$ is the ℓ_1-norm sparse noise term, while λ_{TV}, λ_2 and λ_3 are the regularization parameters corresponding to total variation (TV), joint-sparsity, and sparse noise terms, respectively. In (5), $\text{TV}(\boldsymbol{X})$ is used to promote smoothness or piecewise constant transitions between neighboring pixels in the same endmember.

However, varying noise levels in each band are ignored in traditional SU analysis. SU-NLE [36] and SUBM [37] indicate that each of the spectral dimensions is usually degraded by different noise levels. SUBM starts by estimating the noise weighting matrix $\boldsymbol{W}$ using multiple regression. After that, its unmixing problem is formulated as follows:

$$\min_{\boldsymbol{X},\boldsymbol{E}} \frac{1}{2}\|\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{AX} - \boldsymbol{E})\|_F^2 + \lambda_1\|\boldsymbol{X}\|_{2,1} + \lambda_2\|\boldsymbol{E}\|_1$$
$$\text{s.t. } \boldsymbol{X} \geq 0 \tag{6}$$

where $\boldsymbol{W} \in \mathbb{R}^{L \times L}$ is the diagonal matrix, and $\boldsymbol{W}(i,i) = 1/\delta_i$ for $i \in [1, L]$. Here, δ_i refers to the normalized standard deviation in the ith band.

Next to the robust loss function, latterly, RSU [38] exploits the $\ell_{2,1}$-norm to better handle non-Gaussian noise, where the $\ell_{2,1}$-norm loss function is defined as follows:

$$\|\boldsymbol{AX} - \boldsymbol{Y}\|_{2,1} = \sum_{i=1}^L \left(\sum_{j=1}^P \left((\boldsymbol{AX})_{i,j} - \boldsymbol{Y}_{i,j}\right)^2\right)^{1/2}. \tag{7}$$

With the integration of F-norm and $\ell_{2,1}$-norm, an adaptive σ-norm is proposed [39], which mitigates the drawbacks of F-norm and $\ell_{2,1}$-norm and combines their strengths. Its definition is

$$\|\boldsymbol{AX} - \boldsymbol{Y}\|_\sigma = \sum_{i=1}^L \frac{(1+\sigma)\left\|(\boldsymbol{AX})_{i,:} - \boldsymbol{Y}_{i,:}\right\|_2^2}{\left\|(\boldsymbol{AX})_{i,:} - \boldsymbol{Y}_{i,:}\right\|_2 + \sigma} \tag{8}$$

where $(\boldsymbol{AX})_{i,:}$ and $\boldsymbol{Y}_{i,:}$ refer to the row vectors in the ith band, while $\sigma > 0$ controls the tradeoff between the F-norm and $\ell_{2,1}$-norm. The value of $\|\boldsymbol{AX}-\boldsymbol{Y}\|_\sigma$ tends toward $\|\boldsymbol{AX}-\boldsymbol{Y}\|_{2,1}$ as $\sigma \to 0$, while it behaves more like $\|\boldsymbol{AX}-\boldsymbol{Y}\|_F^2$ as $\sigma \to \infty$.

Last but not least, the maximum correntropy is a popular optimization criterion used in robust signal processing [40], [41]. In [42], the correntropy-based unmixing problem is formulated as follows:

$$\min_{\boldsymbol{X}} \mathcal{C}(\boldsymbol{X}) = -\sum_{i=1}^L \exp\left(\frac{-1}{2\kappa^2}\left\|(\boldsymbol{AX})_{i,:} - \boldsymbol{Y}_{i,:}\right\|_2^2\right)$$
$$\text{s.t. } \boldsymbol{X} \geq 0 \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{x}_j = 1, \quad \text{for} \quad j \in [1, P] \tag{9}$$
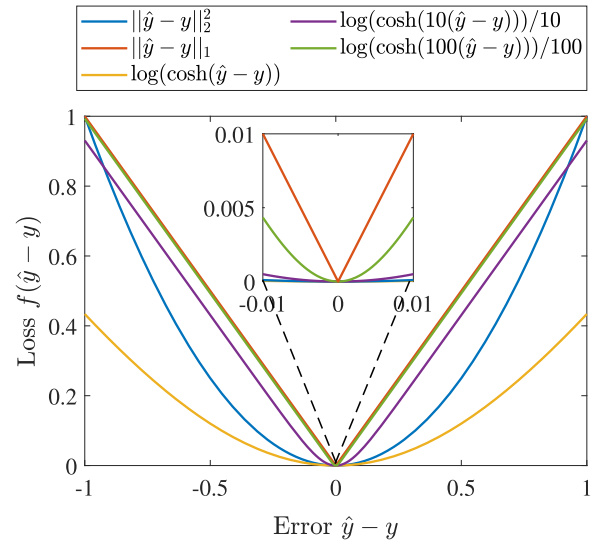

Fig. 1. Different loss functions.

where $\mathcal{C}(\boldsymbol{X})$ represents the correntropy of the Gaussian kernel, and κ denotes the kernel bandwidth which controls its sensitivity, i.e., a small value κ punishes harder the large magnitudes of residual error.

Upon reviewing the above existing works, we notice potential improvements. The selection of regularization parameters, including the sparsity term of $\boldsymbol{X}$ and the sparsity noise term of $\boldsymbol{E}$, is nontrivial and problem-dependent. The efficacy of unmixing is sensitive to the selection of these values. Meanwhile, a robust loss function is preferable in coping with mixed noise. It is worth mentioning that the alternating direction method of multipliers (ADMMs) is prevalently employed in SU problems, such as CSUnL0, SUBM, and RSU. Instead of ADMM, we propose an algorithm with a simpler structure.

III. SU Using ℓ_0-Norm Constraint and Log-Cosh Loss via Projected Gradient Descent

A. Log-Cosh Loss Function

To improve robustness against non-Gaussian noise, we introduce the log-hyperbolic cosine (log-cosh) loss function. We provide a brief overview of the log-cosh features for completeness. The log-cosh loss function belongs to the class of robust estimators. It has been applied in machine learning [43], which includes regression analysis [44], semantic segmentation [45], and variational autoencoder [46]. An elementwise reconstruction loss based on the log-cosh function is given by

$$\frac{\log\left(\cosh\left(a(\hat{y} - y)\right)\right)}{a} \tag{10}$$

where log is the natural logarithm and $a \in \mathbb{R}_+$ is a hyperparameter controlling its sensitivity.

Fig. 1 shows the characteristics of the log-cosh loss function. We observe that the log-cosh loss approaches the ℓ_1-loss as the hyperparameter a increases sufficiently (i.e., $1 \to 100$). More precisely, let $\epsilon = \hat{y} - y$, the first derivative of (10) is $\tanh(a\epsilon)$. Moreover, the first derivative
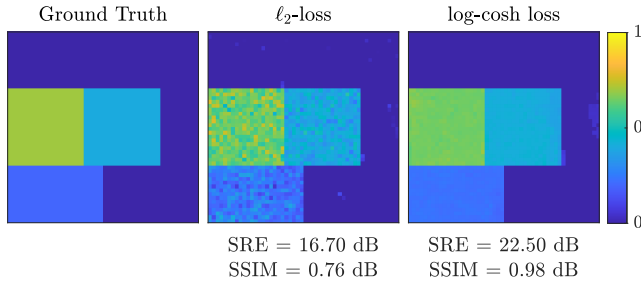
Fig. 2. Abundance estimations by ℓ_2-loss and log-cosh loss under Laplacian noise.

Dataset	Size of image		L-bands	K-endmembers	Size of library
	# height	# width			
DC1	75	75	224	5	224 × 240
DC2	100	100	224	9	224 × 240
Samson	95	95	156	3	156 × 105
Jasper Ridge	100	100	198	4	198 × 529
Urban	307	307	162	6	162 × 651
Cuprite	250	191	188	12	188 × 498

of ℓ_1-loss is the sign($\cdot$) function. As $a \to 100$, $\tanh(100\epsilon)$ approximates the sign($\cdot$) function. Compared with ℓ_2-loss with the gradient equals 2ϵ, the gradient of ℓ_1-loss is 1 with $\epsilon > 0$, and -1 with $\epsilon < 0$, and thus, it has better resistance toward contamination than the ℓ_2-loss [47]. For (10), it is equivalent to $(1/a)\log((e^{a\epsilon} + e^{-a\epsilon})/2)$. When $|\epsilon| \to \infty$, we can express (10) as $|\epsilon| - (1/a)\log 2$ [48]. Apparently, when there are outliers and/or heavy noise, the log-cosh loss behaves like the ℓ_1-loss except for the constant term. In the enlarged part of Fig. 1, log-cosh with $a = 100$ remains smooth in the vicinity of zero, in contrast to the nonsmooth ℓ_1-loss. This implies that the log-cosh loss is continuously differentiable, making it simple to optimize. While as $\epsilon \to 0$, from Taylor expansion, (10) can be approximated as $0.5a\epsilon^2$. Therefore, the log-cosh loss behaves like the ℓ_2-loss close to the origin. In summary, the log-cosh loss is less susceptible to outliers than the ℓ_2-loss, in which the ℓ_2-loss is sensitive to larger errors. To examine its robustness, we have conducted a simple experiment on synthetic data. The synthetic data include five abundance maps of 50×50 pixels with a fixed fraction over a region. Detailed information on the synthetic data are provided in [26]. Laplacian noise has heavy tails to produce extreme values and thus is adopted. We have added zero-mean Laplacian noise to each band whose standard derivation is uniformly distributed within the interval of $[0, 0.5]$. In Fig. 2, it is obvious that the abundance estimation with optimizing log-cosh loss is less affected by the Laplacian noise.

Thereby, we apply the log-cosh loss as the objective function that combines the smoothness of ℓ_2-loss and the robustness of ℓ_1-loss. The convergence of SUNNING can also be analyzed because the log-cosh loss function is convex and L-smooth.

It is worth mentioning that we are aware of other smooth approximations to the ℓ_1-loss. For instance, the Huber loss is a common alternative that has been widely studied [49], [50]. Nevertheless, the Huber loss necessitates adjusting a hyperparameter, whose selection affects performance. Note also that a classical Huber loss function does not have a continuous second derivative.

B. Proposed SU Model

In this section, we formulate the SU problem as follows:

$$\min_{\mathbf{X}} \mathcal{G}(\mathbf{X}) := \frac{\sum_i \sum_j \left(\log\left(\cosh\left(a\left((\mathbf{AX})_{i,j} - \mathbf{Y}_{i,j} \right) \right) \right) \right)}{a}$$
$$\text{s.t. } \mathbf{X} \geq 0, \quad \mathbf{1}^\top \mathbf{X} = \mathbf{1}^\top, \quad \mathbf{X} \text{ is sparse.} \quad (11)$$

Based on (1), (11) can be separated into P-vector subproblems

$$\min_{\mathbf{x}_j} \mathcal{G}(\mathbf{x}_j) := \frac{\sum_{i=1}^{L} \left(\log\left(\cosh\left(a\left(\mathbf{Ax}_j - \mathbf{y}_j \right) \right) \right) \right)}{a}$$
$$\text{s.t. } \mathbf{x}_j \geq 0, \quad \mathbf{1}^\top \mathbf{x}_j = 1, \quad \|\mathbf{x}_j\|_0 \leq S \quad (12)$$

where j corresponds to the jth column of $\mathbf{X}$ and $\mathbf{Y}$. In (12), S indicates the maximum number of active materials allowed in LMM. Unfortunately, from $\|\mathbf{x}_j\|_0 \leq S$, it is hard to determine the exact S in each abundance vector $\mathbf{x}_j$. We roughly set S as the upper bounded sparsity level over the whole HSI scenario. In this work, the studied HSI scenes are known, and we roughly set $S = K$, mentioned in Table I. For unknown HSI data, the number of endmembers, K, could be estimated by the existing methods [13], [51], [52]. Although the parameter S is relaxed to an upper bound level, minimizing log-cosh loss function subject to ANC, ASC, and ℓ_0-norm constraints, have prompted promising unmixing results in our experiments.

C. SUNNING Algorithm

Inspired by the iterative hard thresholding algorithm that handles the ℓ_0-norm sparse recovery problem utilizing the proximal projection approach [53], [54], we adopt the projected gradient descent (PGD) method to solve (12). The two-step alternating scheme is described as follows:

$$\mathbf{z}_j^{(t+1)} = \mathbf{x}_j^{(t)} - \eta \nabla \mathcal{G}\left(\mathbf{x}_j^{(t)} \right) \quad (13a)$$

$$\mathbf{x}_j^{(t+1)} = \mathcal{P}_{\mathcal{C}}\left(\mathbf{z}_j^{(t+1)} \right) \quad (13b)$$

where $\nabla \mathcal{G}(\mathbf{x}_j^{(t)}) = \mathbf{A}^\top \tanh(a(\mathbf{Ax}_j^{(t)} - \mathbf{y}_j))$ is the gradient, $\eta > 0$ is the step size, and $\mathcal{C}$ is the constraint set comprising ANC, ASC, and upper bounded ℓ_0-norm.

For the sake of notational simplicity, we denote $\mathbf{x} = \mathbf{x}_j^{(t+1)}$ and $\mathbf{z} = \mathbf{z}_j^{(t+1)}$. The projection operator $\mathcal{P}_{\mathcal{C}}(\mathbf{z})$ is defined as follows:

$$\mathcal{P}_{\mathcal{C}}(\mathbf{z}) := \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{x}\|_2^2,$$
$$\text{s.t. } \mathbf{x} \geq 0, \quad \mathbf{1}^\top \mathbf{x} = 1, \quad \|\mathbf{x}\|_0 \leq S. \quad (14)$$

The details of $\mathcal{P}_{\mathcal{C}}(\mathbf{z})$ will be discussed in Section III-D.

D. Algorithm for Projection $\mathcal{P}_{\mathcal{C}}(z)$

On the basis of KKT conditions, an efficient nonconvex projection solver is then developed. Without loss of generality, we presume that z is sorted in descending order. Without the ℓ_0-norm constraint, (14) is a projection onto the probabilistic simplex [55], and the Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \mu, \boldsymbol{v}) = \|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + \mu(\mathbf{1}^\top \boldsymbol{x} - 1) - \boldsymbol{v}^\top \boldsymbol{x} \qquad (15)$$

where $\mu \in \mathbb{R}$ is a Lagrange multiplier and $\boldsymbol{v} \in \mathbb{R}_+^n$ is a vector of nonnegative Lagrange multipliers. Due to the convexity, optimal KKT conditions are necessary and sufficient. Based on KKT conditions, we have: 1) stationarity; 2) complementary slackness; 3) primal feasibility; and 4) dual feasibility conditions

$$1) \quad x_i^* = \frac{1}{2}\left(v_i^* - \left(\mu^* - 2z_i\right)\right) \qquad (16a)$$

$$2) \quad v_i^* x_i^* = 0 \qquad (16b)$$

$$3) \quad x_i^* \geq 0, \quad \sum_i^n x_i^* = 1 \qquad (16c)$$

$$4) \quad v_i^* \geq 0 \qquad (16d)$$

for $i \in [1, n]$. Therefore, we identify three cases of $\mu^* - 2z_i$.

1) $\mu^* - 2z_i > 0$: Based on 3), it must hold that $v_i^* \geq (\mu^* - 2z_i) > 0$ to ensure $x_i^* \geq 0$. According to 2), if $v_i^* > 0$ then $x_i^* = 0$ is necessary.

2) $\mu^* - 2z_i = 0$: Based on 2), the sole solution in such a condition is $v_i^* = x_i^* = 0$.

3) $\mu^* - 2z_i < 0$: Based on 4), it must hold that $x_i^* > 0$ since $v_i^* \geq 0$. Additionally, $x_i^* > 0$ implies that $v_i^* = 0$ because of 2). Thus, we obtain $x_i^* = z_i - \mu^*/2$.

To summarize the above results, we have a more compact form

$$x_i^* = \begin{cases} 0, & \text{if } \mu^* - 2z_i \geq 0 \\ z_i - \dfrac{\mu^*}{2}, & \text{if } \mu^* - 2z_i < 0 \end{cases} \qquad (17)$$

for $i \in [1, n]$.

For a given μ^*, only x_i^*'s correspond to z_i's are not 0. Consider $j < i$, if $x_i^* > 0$, then $x_j^* > 0$ according to the descending order of z. Next, we need to compute the optimal value of μ^*. Given that $\sum_i^n x_i^* = 1$, $\sum_i^n x_i^*$ is modified by using (17), thus we have $\mu^* = 2(\sum_{i=1}^n z_i - 1)/n$.

Moving back to (14), we now deal with the ℓ_0-norm constraint. We assume that $s \leq S$, x_i^*'s are positive for $i \in [1, s]$, and x_i^*'s are 0 for $i \in [s + 1, n]$. Followed by condition 3), $\sum_{i=1}^s x_i^* = 1$. With appropriate term rearrangements, we have

$$\mu_s^* = \frac{2\left(\sum_{i=1}^s z_i - 1\right)}{s}. \qquad (18)$$

An iterative scheme is designed to find the optimal element set $\boldsymbol{x}^*$. Starting with $s = 1$, we calculate $\mu_{s=1}^*$ using (18), then we verify the conditions based on (17). After that, s would be increased, and we perform the previous steps again until $S = s$, or the iteration would be stopped at the condition $z_s \leq \mu_s^*/2$. A more efficient way is to check $z_{s+1} \leq \mu_s^*/2$,

avoiding the redundant restoration of the previous status if the condition $z_s \leq \mu_s^*/2$ occurs at the sth iteration.

The procedure for nonconvex projection $\mathcal{P}_{\mathcal{C}}(z)$ is outlined in Algorithm 1. We improve the calculation of μ_s^* to save the computation during each iterative update, $(s-1)\mu_{s-1}^* + 2z_s$ is a more efficient substitute for $2(\sum_{i=1}^s z_i - 1)$.

Algorithm 1 Projection Operator $\mathcal{P}_{\mathcal{C}}(z)$

1 Input: z and S.
2 Initialization:
3 Sort z in descending order to get $\tilde{z}$.
4 Set $s = 1$, $\mu_1 = 2\tilde{z}_1 - 2$, and $\tilde{x} = \boldsymbol{0}$.
5 if $\tilde{z}_2 \leq \dfrac{\mu_1}{2}$ or $S = 1$ **then**
6 $\quad \tilde{x}_1 = \tilde{z}_1 - \dfrac{\mu_1}{2}$.
7 $\quad$ Jump to Step 18.
8 end
9 for $s = 2, \ldots, S - 1$ **do**
10 $\quad \mu_s = \dfrac{(s-1)\mu_{s-1} + 2\tilde{z}_s}{s}$,
11 $\quad$ **if** $\tilde{z}_{s+1} \leq \dfrac{\mu_s}{2}$ **then**
12 $\quad\quad \tilde{x}_i = \tilde{z}_i - \dfrac{\mu_s}{2}$ for $i \in [1, s]$.
13 $\quad\quad$ Jump to Step 18.
14 $\quad$ **end**
15 end
16 $\mu_S = \dfrac{(S-1)\mu_{S-1} + 2\tilde{z}_S}{S}$.
17 $\tilde{x}_i = \tilde{z}_i - \dfrac{\mu_S}{2}$ for $i \in [1, S]$.
18 Map $\tilde{x}$ into the original ordering to get $\boldsymbol{x}^*$.
19 Output: $\boldsymbol{x}^*$.

Algorithm 2 SUNNING for Solving (11)

1 Input: $\boldsymbol{A}$, $\boldsymbol{Y}$, a, S, α and $t_{\max}$.
2 Initialization:
3 Initialize $\boldsymbol{X}^{(0)}$ and $\eta = \alpha/L$.
4 for $t = 1, \ldots, t_{\max}$ **do**
5 $\quad$ Calculate $\nabla \mathcal{G}(\boldsymbol{X}^{(t)}) = \boldsymbol{A}^\top \tanh(a(\boldsymbol{A}\boldsymbol{X}^{(t)} - \boldsymbol{Y}))$.
6 $\quad$ Update $\boldsymbol{Z}^{(t+1)} = \boldsymbol{X}^{(t)} - \eta \nabla \mathcal{G}(\boldsymbol{X}^{(t)})$.
7 $\quad$ **for** $j = 1, \ldots, P\text{–}pixels$ **do in parallel**
8 $\quad\quad$ Extract $\boldsymbol{z}_j^{(t+1)}$ from the j-th column of $\boldsymbol{Z}^{(t+1)}$.
9 $\quad\quad$ Compute $\boldsymbol{x}_j^{(t+1)} = \mathcal{P}_{\mathcal{C}}(\boldsymbol{z}_j^{(t+1)})$ via Algorithm 1.
10 $\quad$ **end**
11 $\quad$ **Stop** if $t = t_{\max}$ or (19) is satisfied.
12 end
13 Output: $\boldsymbol{X}$.

Algorithm 2 provides the pseudocode of SUNNING. There are four parameters, including hyperparameter a, sparsity S, step-size η, and maximum iteration number $t_{\max}$. To approximate the ℓ_1-loss, $a = 100$ is employed in all experiments. S is set to the number of endmembers desired. For choosing a learning rate η, $\eta \in (0, \alpha/L]$ and $0 < \alpha \leq 1$ are commonly used [56]. The Lipschitz constant L
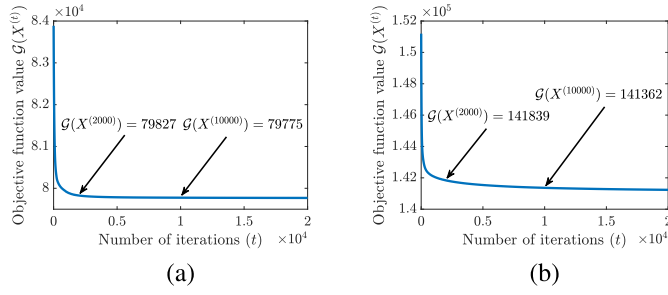
Fig. 3. Convergence behavior of SUNNING. (a) DC1. (b) DC2.

Fig. 4. Illustration of SLP strategy.

is the largest eigenvalue of $\nabla^2 \mathcal{G}(\boldsymbol{x}_j^{(t)})$. Note that $\nabla^2 \mathcal{G}(\boldsymbol{x}_j^{(t)}) = a\boldsymbol{A}^\top \text{diag}(1 - \tanh^2(a(\boldsymbol{A}\boldsymbol{x}_j^{(t)} - \boldsymbol{y}_j^{(t)})))\boldsymbol{A}$, where $\nabla^2 \mathcal{G}(\boldsymbol{x}_j^{(t)})$ is the positive semidefinite. Details will be presented in Appendix A.

In SUNNING, there are two stopping criteria. The first one is to terminate at $t = t_{\max}$. The objective function value in the experiments shows that $t_{\max} \geq 2000$ is sufficient to ensure convergence. Optionally, the algorithm could be stopped by the following criterion [57]:

$$\frac{\|\boldsymbol{x}_j^{(t+1)} - \boldsymbol{x}_j^{(t)}\|_2^2}{S} \leq 10^{-8} \quad \text{for} \quad j \in [1, P]. \quad (19)$$

E. Convergence Analysis of SUNNING

In SUNNING, it solves a total of P subproblems, and each subproblem is solved by the same procedure. To prove the convergence, we theoretically analyze the sequence of the objective function value. Theorem 1 establishes the convergence behavior of the objective function value.

Theorem 1: In SUNNING, the objective function value $\mathcal{G}(\boldsymbol{x}) = \sum_{i=1}^{L}(\log(\cosh(a(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}))))/a$ for $\boldsymbol{x} \in \mathbb{R}^n$ satisfies the following properties.
1) $\mathcal{G}(\boldsymbol{x})$ is the lower bounded.
2) $\nabla \mathcal{G}(\boldsymbol{x})$ is the Lipschitz continuous.
3) If $0 < \eta < 1/L$, then $\mathcal{G}(\boldsymbol{x}^{(t+1)}) \leq \mathcal{G}(\boldsymbol{x}^{(t)})$, where L is the L-Lipschitz constant and the equal sign holds if and only if $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)}$.

Therefore, the convergence of $\{\mathcal{G}(\boldsymbol{x}^{(t)})\}_{t=0}^{\infty}$ can be guaranteed with $0 < \eta < 1/L$. The proof is provided in Appendix A.

In addition, we analyze the convergence behavior of variable sequence $\{\boldsymbol{x}^{(t)}\}$ in Theorem 2.

Theorem 2: With the initialization of $\|\boldsymbol{x}^{(0)}\|_2^2 < +\infty$ and $\mathcal{G}(\boldsymbol{x}^{(0)}) < +\infty$. The variable sequence $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$ generated by SUNNING has the following properties.
1) $\mathcal{G}(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\|_2 \to \infty$.
2) $\boldsymbol{x}^{(t+1)}$ is bounded.
3) $\lim_{t\to\infty} \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 = 0$ as $\mathcal{G}^* = \inf \mathcal{G}(\boldsymbol{x})$ is finite.

Based on $\mathcal{G}(\boldsymbol{x}^{(t+1)}) < \mathcal{G}(\boldsymbol{x}^{(t)})$ with $\boldsymbol{x}^{(t+1)} \neq \boldsymbol{x}^{(t)}$, we guarantee that variable sequence converges to a solution. The proof is provided in Appendix B.

Fig. 3 shows the convergence behavior of SUNNING on synthetic datasets. It is apparent that the objective function value $\mathcal{G}(\boldsymbol{X}^{(t)})$ begins to converge at around $t = 2000$.

F. Computational Complexity

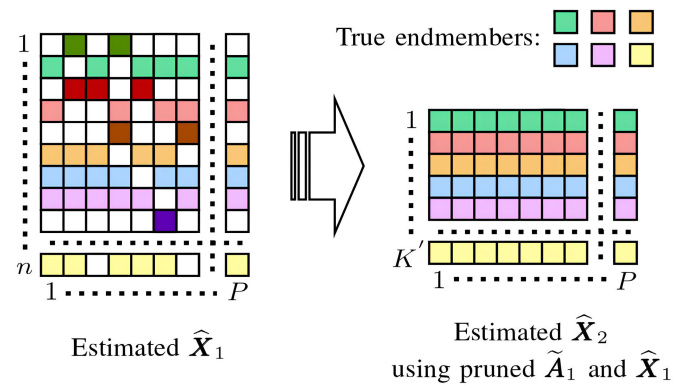In Algorithm 1, the initial step involves computing a temporary solution $\boldsymbol{Z}$ using gradient descent, which has a complexity of $\mathcal{O}(LnP)$. Subsequently, a projection of $\boldsymbol{Z}$ is carried out in parallel. This leads to P projection operations. For each projection, the sorting operation has a complexity of $\mathcal{O}(n \log n)$, and the computation of μ can take up to a maximum of S attempts. Consequently, the complexity of a single projection is $\mathcal{O}(n \log n + S)$. In summary, SUNNING has an overall complexity of $\mathcal{O}(t_{\max}(LnP + P(n \log n + S)))$.

IV. SU WITH SLP STRATEGY

In the previous section, we discussed estimating each abundance vector $\boldsymbol{x}_j$ independently using each hyperspectral pixel $\boldsymbol{y}_j$ and the spectral library $\boldsymbol{A}$. While the reality is that $\boldsymbol{X}$ has the row sparsity property, we can utilize the abundance results of SUNNING to prune irrelevant endmembers in $\boldsymbol{A}$ and the virtual abundances in $\boldsymbol{X}$. Thus, this section describes SUNNING–SLP, a modified version of SUNNING that applies an effective pruning strategy.

In most cases, the number of endmembers in a given HSI scene is substantially lower than the number of spectral signatures in $\boldsymbol{A}$, i.e., $K \ll n$. Therefore, we are motivated to prune the rudimentary result $\widehat{\boldsymbol{X}}$, i.e., remove the inactive and/or weakly distributed spectral signatures. Existing works [58], [59], [60] demonstrate that the SLP approach can enhance the accuracy of the unmixing task and also reduce the complexity of the SU algorithm. In SUNNING, if $\boldsymbol{y}_j$ is heavily corrupted by mixed noise, then $\boldsymbol{x}_j$ estimates the abundance values poorly as it tends toward fitting the noisy $\boldsymbol{y}_j$. As depicted in Fig. 4, $\widehat{\boldsymbol{X}}_1$ is the estimated abundance matrix, wherein some abundance vectors $\boldsymbol{x}_j$'s unmix out nonexistent endmembers. We assume that nonexistent endmembers are predominantly inactive and/or sparsely distributed in practice. Based on the contribution of the endmembers in $\widehat{\boldsymbol{X}}_1$, the less significant endmembers would be eliminated. Then, the abundance matrix $\widehat{\boldsymbol{X}}_2$ is re-estimated using subdictionary $\widetilde{\boldsymbol{A}}_1$. We progressively update $\widehat{\boldsymbol{X}}$ by repeating dictionary pruning and re-estimation, and finally, a fine-quality abundance matrix can be obtained. Similar to [58], SLP strategy is broken down into two main steps.
1) Every row vector in $\widehat{\boldsymbol{X}}$ represents a unique endmember. We compute the number of pixels for each endmember whose abundance value is below a predefined threshold, denoted by ϕ.

Algorithm 3 SLP Strategy

---

1 **Initialization:**
2 Set $q = 1$, ϕ, Δ, K, and $D_1 = \text{size}(\widetilde{\boldsymbol{A}}_1, 2) - K$.
3 **while** $D_q \geq \Delta$ **do**
4   Obtain $\widehat{\boldsymbol{X}}_q$ via Algorithm 2.
5   **for** $i = 1, \ldots, \text{size}(\widetilde{\boldsymbol{A}}_q, 2)$ **do**
6   Set $C_i = 0$.
7   **for** $j = 1, \ldots, P\text{–pixels}$ **do**
8   **if** $\widehat{\boldsymbol{X}}_q(i, j) < q \times \phi$ **then**
9                 $C_i = C_i + 1$.
10   **end**
11   **end**
12   **if** $C_i = P\text{–pixels}$ **then**
13   Prune $\widehat{\boldsymbol{X}}_{[i,:]}$ and $\widetilde{\boldsymbol{A}}_{[:,i]}$.
14   **end**
15   **end**
16     $\widetilde{\boldsymbol{A}}_{q+1} = \widetilde{\boldsymbol{A}}_q$.
17   Update $D_{q+1} = \text{size}(\widetilde{\boldsymbol{A}}_{q+1}, 2) - K$.
18     $q = q + 1$.
19   **Stop** if $D_{q-1} = 0$.
20 **end**
21 Restore $\widehat{\boldsymbol{X}}$ into the original space $\mathbb{R}^{n \times P}$.
22 **Output:** $\widehat{\boldsymbol{X}}$.

---

  2) If the number of pixels counted from 1) is equal to a total of P pixels, the corresponding ith endmember will be removed from $\widehat{\boldsymbol{X}}$, which will result in the removal of a row vector $\widehat{\boldsymbol{X}}_{[i,:]}$. Again, the corresponding column $\boldsymbol{A}_{[:,i]}$ will be discarded from $\boldsymbol{A}$.

The key rationale behind the SLP strategy is that the dropped ith endmember could be considered inactive or nongenuine if the abundance value is below a lower bound, i.e., rarely distributed across the HSI scene. In general, if a spectral signature corresponds to a genuine endmember, its abundance in certain regions should be highly active. This behavior serves as the basis for trimming the spectral library $\boldsymbol{A}$.

Algorithm 3 summarizes the steps for performing an SLP strategy integrated with our SUNNING. In Algorithm 3, $q \times \phi$ refers to an iterative threshold representing the minimum abundance value, where q is iteration number and $\phi > 0$ is the preset threshold. Algorithm 3 terminates if the difference D_{q+1} between the number of preserved spectral signatures, $\text{size}(\widetilde{\boldsymbol{A}}_{q+1}, 2)$, and the total number of endmembers, K, is less than Δ, or $D_{q-1} = 0$. If $\Delta = 10$, the preserved spectral signatures in $\widetilde{\boldsymbol{A}}_{q+1}$ will not be more than $K + 10$. The ϕ is usually initialized with a small value, whereas Δ could be set to a large value or equal to 0. For $\Delta = 0$, it is a hard estimation that only uses K signatures to determine $\widehat{\boldsymbol{X}}_{q+1}$. For $\Delta > 0$, it signifies to a soft estimation that more signatures are retained for estimating $\widehat{\boldsymbol{X}}_{q+1}$, preventing the removal of some genuine endmembers from $\widetilde{\boldsymbol{A}}_q$. For the settings of Δ and ϕ, extensive experiments are conducted in [58] to analyze the parameter settings of ϕ and Δ in relation to signal-to-reconstruction error (SRE).

Following previous works, we set $\phi = 0.02$ and $\Delta = 1$ in the synthetic experiments. For the value of K, it is provided in Table I. Additionally, an experiment on Δ is conducted which will be discussed in Section V-G.

V. EXPERIMENTAL RESULTS

In this section, we compare our approach with 12 SU methods, namely, JSTV [26], MUA–BPT [61], LRUnSAL–TV [62], CSUnL0–ADL [39], SUBM–ADL [39], SUnSAL–ADL [39], Pro–A–BM3D [63], LGSU [64], MdLRR [65], LSU [66], MIHT [57], and RHUIDR–HTV [67]. It is worth noting that CSUnL0–ADL, SUBM–ADL, and SUnSAL–ADL are the extension versions of CSUnL0 [31], SUBM [37], and SUnSAL [32], respectively. In [39], the adaptive loss function is applied to these conventional SU methods to achieve robustness. In parameter settings, we test different values among the regularization parameters in each of the regularized-based SU methods, i.e., [1e−4, 1e−3, ..., 10, 100]. Then, the parameters with the highest SRE are selected. To provide a fair comparison, SUNNING–SLP is only used to verify the effectiveness of SLP by comparing it with SUNNING.

A. Performance Metrics

In performance evaluation, the SRE and the mean structural similarity index measure (MSSIM) [68] are used. The SRE in decibel (dB) is defined as follows:

$$\text{SRE} = 10 \log_{10} \frac{\|\boldsymbol{X}\|_F^2}{\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_F^2} \tag{20}$$

where $\boldsymbol{X}$ denotes the ground-truth abundance and $\widehat{\boldsymbol{X}}$ denotes the estimate. A large SRE indicates a highly accurate reconstruction, and vice versa.

The abundance map has image-like characteristics. To better illustrate the similarity between ground-truth abundance and estimated abundance in terms of structure, contrast, and luminance, MSSIM is adopted. First, the estimated abundance $\widehat{\boldsymbol{X}}$ is converted from an $n \times P$ matrix to a height $\times$ width $\times n$ abundance cube. After that, the corresponding K-endmembers that are expected to be true are extracted from the abundance cube, namely, $\widehat{\boldsymbol{X}}_c$. The MSSIM is defined as follows:

$$\text{MSSIM} = \frac{1}{K} \sum_{i=1}^{K} \text{SSIM}\left(\boldsymbol{X}_c(:,:,i), \widehat{\boldsymbol{X}}_c(:,:,i)\right) \tag{21}$$

where $\boldsymbol{X}_c(:,:,i)$ and $\widehat{\boldsymbol{X}}_c(:,:,i)$ represent the corresponding reference abundance map and the estimated abundance map, respectively. Here, $\text{SSIM}(\cdot, \cdot)$ calculates the structural similarity index according to [68].

B. Simulation Configurations

Six HSI datasets are used to evaluate the performance, including two synthetic datasets and four real-world datasets. Their details are summarized in Table I.

We first consider a synthetic library $\boldsymbol{A}_{\text{USGS}} \in \mathbb{R}^{224 \times 240}$ generated by selecting a subset of 240 materials from the

United States Geological Survey (USGS) library.[1] In $\boldsymbol{A}_{\mathrm{USGS}}$, the angle between any pair of spectral signatures is at least 4.44°, and each spectral signature has 224 bands with wavelengths ranging from 0.38 to 2.5 μm.

The synthetic data cube 1 (DC1) [22] is a widely used dataset, which consists of 224 bands and 75×75 pixels with five endmembers from $\boldsymbol{A}_{\mathrm{USGS}}$. These selected endmembers are: #1 Jarosite, #2 Calcite, #3 Howlite, #4 Fassaite, and #5 Andradite. It comprises the pure and mixed regions, in which the fractional abundances of these endmembers are spatially distributed as square regions. Each mixed region is composed of 2–5 endmember composites, whereas the fractional abundances of five endmembers in background pixels are fixed to [0.1149, 0.0741, 0.2003, 0.2055, 0.4051], respectively.

The synthetic data cube 2 (DC2) [60], [69] is also frequently used. It consists of nine endmembers from $\boldsymbol{A}_{\mathrm{USGS}}$ and contains 100×100 pixels. These selected endmembers are: #1 Jarosite, #2 Calcite, #3 Howlite, #4 Fassaite, #5 Andradite, #6 Hypersthene, #7 Opal, #8 Nacrite, and #9 Sepiolite. Its abundance maps follow the Dirichlet distribution, demonstrating piecewise smooth maps and behaving with steep transitions.

Four real-world HSIs are considered in our work, viz., Samson, Jasper Ridge, Urban, and Cuprite. Since the size of these original data is large, we take into account their subimages, those details are provided in Fig. 5 and Table I. As the ground-truth abundance maps are unavailable to these real HSIs, we compare the estimated results based on intuitive visual observation. Works[2] [70], [71], [72] have provided the reference abundance maps with labels to Samson, Jasper Ridge, and Urban HSIs. Nevertheless, we cannot use these labels for quantitative measurement because the spectral signatures used in our experiments are inconsistent with them. Due to signature variability, unmixing with varying degrees of spectral signatures limits the accuracy and results in modeling errors [12], [73].

In Cuprite, the spectral library $\boldsymbol{A}_{\mathrm{USGS}} \in \mathbb{R}^{188 \times 498}$ is built using 498 spectral signatures in the USGS library after removing the water absorption and low signal-to-noise ratio (SNR) bands [30], [35]. Since there are many types of minerals, we only present three dominant minerals, viz., #1 Alunite, #3 Buddingtonite, and #12 Chalcedony.

Similarly, undesirable and corrupted bands in the Samson, Jasper Ridge, and Urban HSIs have been removed [72], [74]. Given that spectral libraries $\boldsymbol{A}$'s containing true spectra in those scenarios are unavailable, the previous works have been tried to implement an image-based library construction method to extract $\boldsymbol{A}$'s from HSIs [61], [75]. The procedure is to extract pure pixels from an HSI, then pruning insignificant features to limit the number of features for each endmember. Thanks to their contributions, the preliminary spectral libraries[3] for these scenes have been prepared. Notably, the spectral library $\boldsymbol{A}_{\mathrm{HSI}}$ has a total of K subgroups of spectral features. We denote

them as $\boldsymbol{A}_k$ for $k \in [1, K]$. Furthermore, all the spectral features within the kth subgroup belong to the same type of material. After estimating $\widehat{\boldsymbol{X}}$, we divide $\widehat{\boldsymbol{X}}$ into K subgroups and compute the average abundance value for each pixel in each subgroup $\widehat{\boldsymbol{X}}_k$. Subsequently, K row vectors are obtained, i.e., $\widehat{\boldsymbol{X}}_{\mathrm{AVG}} \in \mathbb{R}^{K \times P}$, and then, the final result is normalized to ensure that ASC is satisfied. These operations result in a single set of abundance maps for each underlying endmember.

To validate the robustness of our approach in hyperspectral unmixing, two mixed noise cases, *Noise Case 1* and *Noise Case 2*, have been set up in all the datasets.

1) *Noise Case 1:* Both DC1 and DC2 are corrupted by non- independent identically distributed (i.i.d.) white Gaussian noise. We add noise to each band with the SNR ranging from 20 to 35 dB. Next, uniform noise distributed between [−1, 1] is added to 10% of randomly selected pixels in the entire HSI. For the real data, the existence of Gaussian noise is assumed. Only uniform noise is added to 10% of randomly selected pixels.

2) *Noise Case 2:* Both DC1 and DC2 are contaminated by white Gaussian noise first, following the same steps as *Noise Case 1*. Next, 5% salt-and-pepper is added to the entire HSI. Third, 10% of the spectral bands are selected at random. In each selected band, 10% of horizontal line strips are degraded randomly. Again, 10% bands are selected randomly, and 10% of random vertical line strips are degraded in each corresponding band. The intensity of such horizontal and vertical strips varies between [0, 1]. For the real data, the addition of white Gaussian noise is skipped.

C. Simulation Results With DC1 and DC2

In this section, we investigate the performance of various SU algorithms. The 50 independent trials are conducted in each noise case. Tables II and III show the average SREs and MSSIMs as well as the standard derivations, where DC1 and DC2 are subject to mixed noise contamination.

It is seen from Table II that the SUNNING obtains the highest SRE values in both noise cases. Furthermore, our method is comparable with the robust methods, such as JSTV and LRUnSAL–TV, in terms of MSSIM. For instance, in *Noise case 1*, the SRE of SUNNING is 20.15 ± 0.59 dB, while its MSSIM is 0.88 ± 0.01. Because of piecewise smooth regularization, JSTV, LRUnSAL–TV, and RHUIDR–HTV can characterize the abundance maps well in homogeneous regions, such as background areas. Their solutions get spatial consistency, so the spatial distribution of the materials is estimated with satisfactory accuracy, as listed in Tables II and III. In addition, Fig. 6 presents the estimated abundance maps obtained on DC1 using various algorithms. It is observed that the abundance maps by SUNNING are more precise than those by the competing methods. Especially, the pure and mixed regions in endmembers, #2 Calcite and #4 Fassaite, our SUNNING correctly unmixes their distribution positions and their corresponding proportions of abundance, whereas the robust methods cannot estimate the positions and

---

TABLE II
SREs, MSSIMs, AND STANDARD DERIVATIONS OF 12 ALGORITHMS ON DC1

Noise	Metric	SUNNING	JSTV	MUA–BPT	LRUnSAL–TV	CSUnL0–ADL	SUBM–ADL	SUnSAL–ADL
	SRE	20.15±0.59	15.29±0.45	9.68±0.50	14.23±0.81	3.63±0.46	6.08±0.37	2.73±0.02
	MSSIM	0.88±0.01	0.89±0.006	0.66±0.02	0.87±0.01	0.11±0.02	0.39±0.03	0.10±0.002
Noise Case 1		Pro–A–BM3D	LGSU	MdLRR	LSU	MIHT	RHUIDR–HTV	
	SRE	7.97±0.06	3.43±0.04	4.98±0.09	4.25±0.37	11.25±0.49	11.21±1.30	
	MSSIM	0.29±0.003	0.13±0.002	0.32±0.005	0.17±0.03	0.47±0.04	0.85±0.03	
		SUNNING	JSTV	MUA–BPT	LRUnSAL–TV	CSUnL0–ADL	SUBM–ADL	SUnSAL–ADL
	SRE	20.28±0.57	15.25±0.49	5.84±0.51	14.21±0.74	5.12±1.12	5.54±0.73	3.38±0.06
	MSSIM	0.88±0.01	0.89±0.006	0.48±0.04	0.87±0.01	0.20±0.04	0.49±0.08	0.14±0.003
Noise Case 2		Pro–A–BM3D	LGSU	MdLRR	LSU	MIHT	RHUIDR–HTV	
	SRE	9.62±0.18	4.11±0.08	6.04±0.39	4.97±0.61	6.43±0.95	14.94±0.93	
	MSSIM	0.38±0.005	0.16±0.005	0.54±0.05	0.20±0.05	0.35±0.06	0.83±0.02	

TABLE III
SREs, MSSIMs, AND STANDARD DERIVATIONS OF 12 ALGORITHMS ON DC2

Noise	Metric	SUNNING	JSTV	MUA–BPT	LRUnSAL–TV	CSUnL0–ADL	SUBM–ADL	SUnSAL–ADL
	SRE	18.30±0.36	14.09±0.69	6.94±0.45	17.01±0.29	7.30±1.50	6.03±1.69	2.81±0.03
	MSSIM	0.58±0.01	0.72±0.01	0.41±0.01	0.59±0.01	0.32±0.01	0.51±0.05	0.31±0.002
Noise Case 1		Pro–A–BM3D	LGSU	MdLRR	LSU	MIHT	RHUIDR–HTV	
	SRE	7.36±0.10	5.26±0.16	9.16±0.05	7.20±0.74	8.16±0.48	15.00±1.51	
	MSSIM	0.21±0.003	0.39±0.01	0.46±0.002	0.31±0.01	0.39±0.01	0.67±0.02	
		SUNNING	JSTV	MUA–BPT	LRUnSAL–TV	CSUnL0–ADL	SUBM–ADL	SUnSAL–ADL
	SRE	18.33±0.39	14.24±0.70	3.26±0.40	17.03±0.32	9.17±1.79	8.23±1.84	3.80±0.08
	MSSIM	0.58±0.02	0.72±0.01	0.37±0.01	0.59±0.01	0.36±0.01	0.44±0.03	0.33±0.003
Noise Case 2		Pro–A–BM3D	LGSU	MdLRR	LSU	MIHT	RHUIDR–HTV	
	SRE	9.35±0.16	6.92±0.14	8.74±0.10	8.14±0.44	4.29±0.46	18.76±1.35	
	MSSIM	0.28±0.004	0.37±0.004	0.46±0.004	0.36±0.01	0.34±0.01	0.75±0.02	

proportions of abundance accurately, such as including JSTV, LRUnSAL–TV, and RHUIDR–HTV.

One of the underlying reasons is that they combine multiple objectives into a single objective function, but these penalty terms control the relative importance between the regularization and the original objective. As mentioned in [67] and [76], the importance of original objective loss becomes less significant as more severe penalties are imposed on multiple regularization terms. Also, the row-sparsity term cannot work perfectly in the mixed noise scene. For example, in $\ell_{2,1}$-norm, the regularization might not keep sufficiently sparse, $\lambda \sum_{i=1}^{n} \|\boldsymbol{x}^{i,:}\|_2$ may induce irrelevant spectral signatures $\boldsymbol{x}^i$'s as spatial locations $\boldsymbol{x}_j$'s are affected by mixed noise.

In Fig. 7, we observe that the ground truth on DC2 not only has densely and strongly distributed values in certain regions but is also sparsely and weakly distributed in arbitrary placements. Such abundances are arbitrarily distributed, their abundance values are close to 0, making it complicated to separate out insignificant signatures. In addition, DC2 is a highly mixed scenario with nine endmembers, thereby DC2 with additive mixed noise would make the unmixing

task challenging. According to Table III, SUNNING obtains higher SREs than most of the methods on both noise cases. However, MSSIM is sensitive to distorted abundance maps. The MSSIMs of SUNNING and the competing algorithms are not performing well enough. To visualize this difference, Fig. 7 shows that SUNNING, JSTV, LRUnSAL–TV, and RHUIDR–HTV roughly approximate the abundance maps, while yet the precision of abundance value is lost. Moreover, their abundance maps omit sparse and weak contents, as shown in endmembers, #1 Jarosite, #2 Calcite, and #5 Andradite. To tackle this issue, the SLP approach is considered and we will discuss SUNNING–SLP in Section V-E.

In Table IV, we list the runtime comparison to all methods on synthetic datasets. The associated number of runs on each method is also specified. The runtime is measured using MATLAB with a 3.6-GHz Intel Core i7-12700K. Apparently, both MUA–BPT and Pro–A–BM3D require less computational time, but their unmixing results are worse than most methods. Noticeably, LRUnSAL–TV performs well in terms of time-effectiveness and robustness. While our SUNNING can provide improved robust performance with comparable runtime.
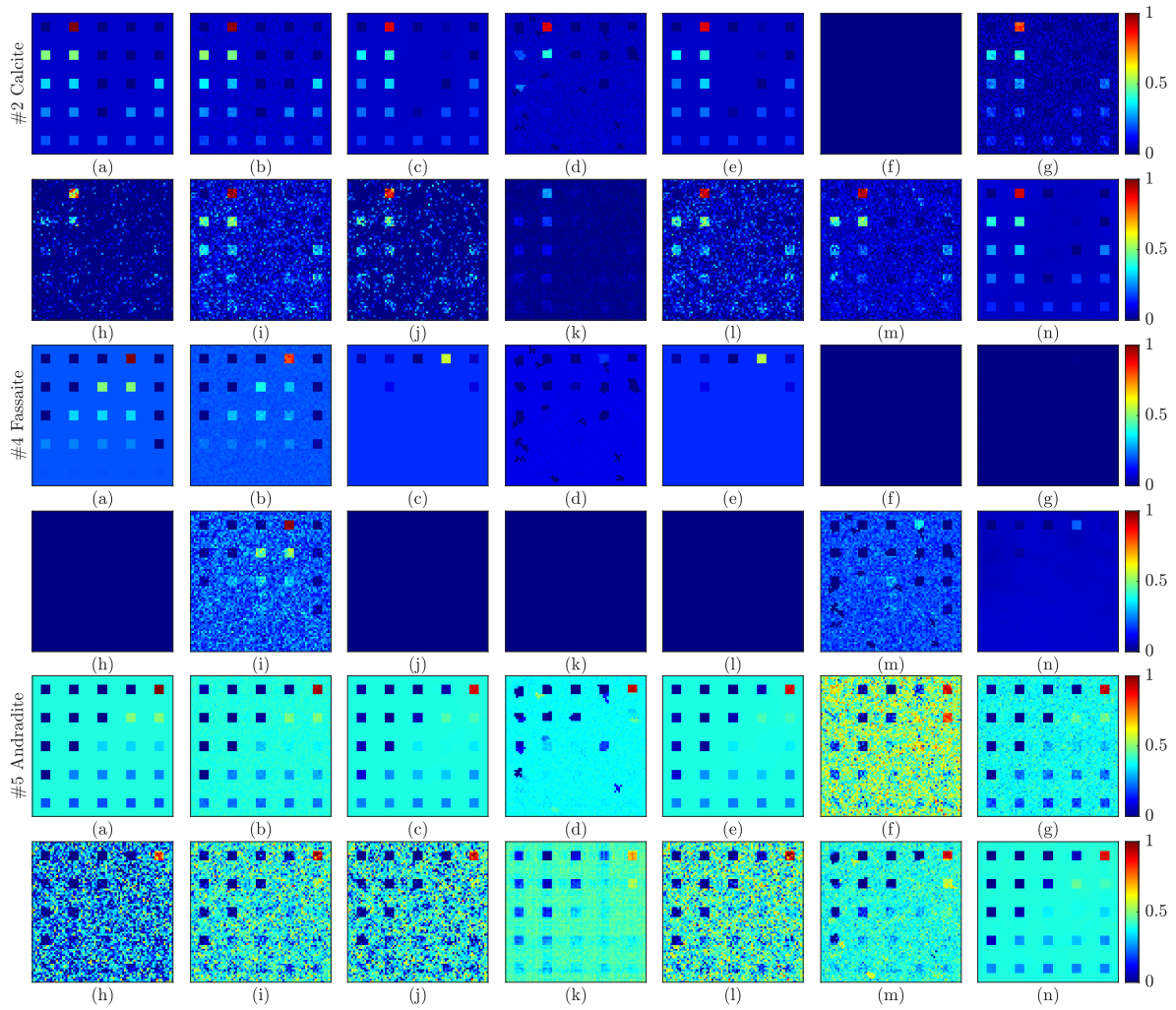
Fig. 5. Ground truth and abundance maps obtained by various algorithms on DC1 under *Noise Case 1*. Three selected abundance maps correspond to endmembers {#2 Calcite, #4 Fassaite, #5 Andradite}. (a) True. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.
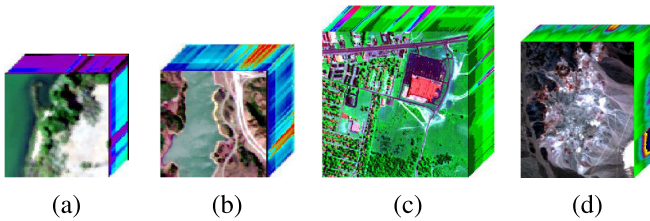

Fig. 6. Illustration of real hyperspectral subimages. (a) Samson. (b) Jasper ridge. (c) Urban. (d) Cuprite.

D. Simulation Results With Real HSI Data

In this section, we further investigate the unmixing performance on real-world data. Due to the page limit, the simulation results for the Samson and Jasper Ridge HSIs are included in the Supplemental Material. Figs. 8 and 9 present the abundance maps estimated by various SU methods on different noise-polluted Urban HSIs. These two figures reveal the unsatisfactory performance of MUA–BPT, CSUnL0–ADL, SUnSAL–ADL, Pro–A–BM3D, LGSU, MdLRR, LSU, and MIHT. Clearly, their abundance estimates are heavily affected by mixed noise. Their distribution positions and proportions of endmembers are biased toward fitting the noise. In comparison, our SUNNING, JSTV, LRUnSAL–TV, SUBM–ADL, and RHUIDR–HTV yield robust visual effects. The composite abundance maps of SUNNING and robust methods are spatially congruent with the labeled reference, where the recovered abundances preserve fine variability and intricate structures.

Although JSTV, LRUnSAL–TV, SUBM–ADL, and RHUIDR–HTV provide robust results in Figs. 8 and 9, they may overestimate the fractional abundance maps in certain regions. For example, the importance of #5 Metal is overestimated in their estimates. We can observe that they provide virtual abundance information in #5 Metal. These regions are supposed to be inactive, compared with the reference ones. It is seen that the abundance maps estimated by SUNNING are closer to the reference ones.

Moreover, our SUNNING improves the unmixing results in Cuprite HSI to some degree. We observe from the zoomed details of Figs. 10 and 11 that the abundance maps estimated by SUNNING are more similar to the mineral distribution maps analyzed by Tricorder 3.3 software.[4] Note

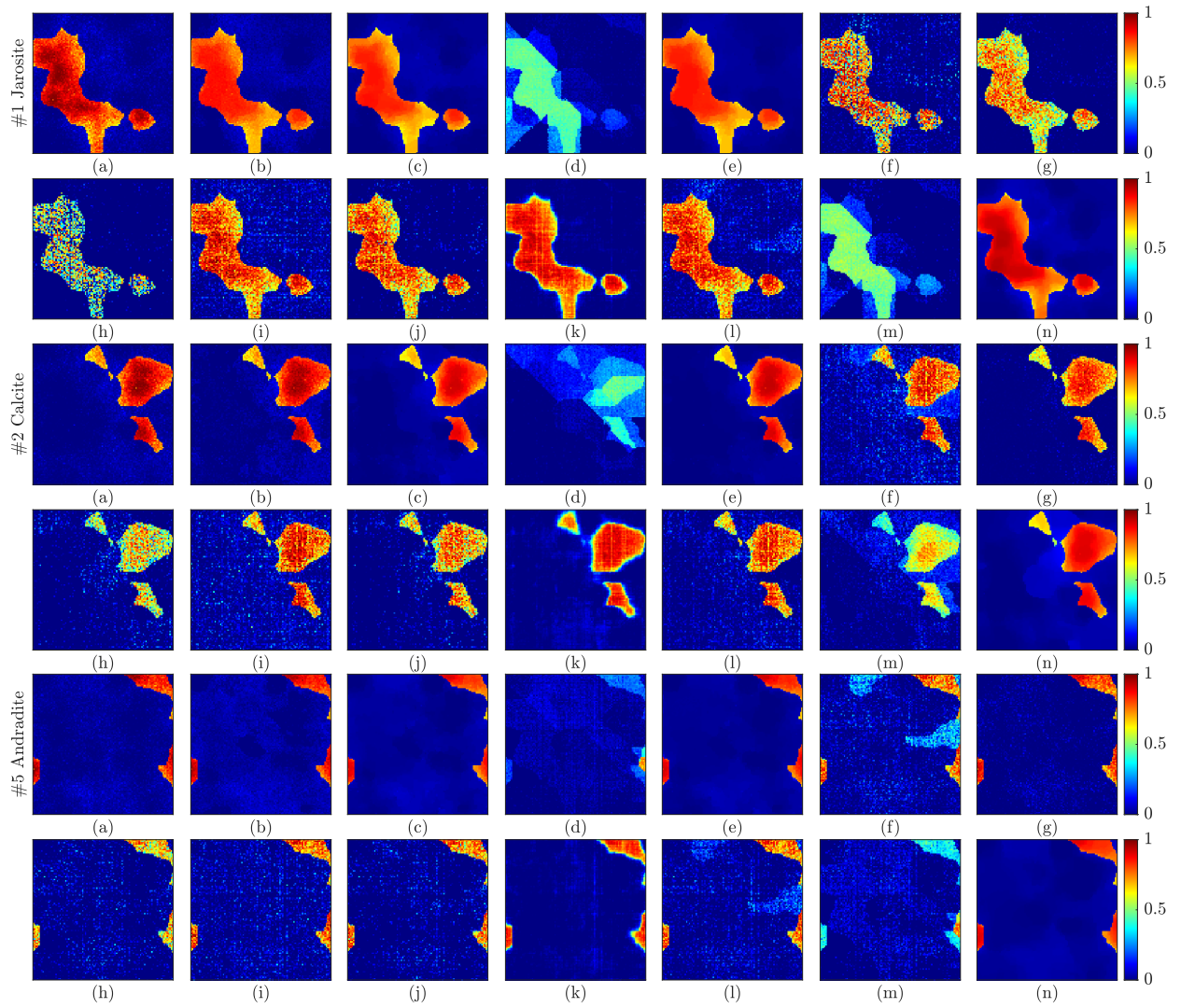[4]Available online at http://speclab.cr.usgs.gov/PAPER/tetracorder

Fig. 7. Ground truth and abundance maps obtained by various algorithms on DC2 under *Noise Case 2*. Three selected abundance maps correspond to endmembers {#1 Jarosite, #2 Calcite, #5 Andradite}. (a) True. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.

that Tricorder 3.3 software is used to classify each pixel into a specific mineral class but without proportional information about each mineral [77], [78]. Under the same intensity range, the abundance maps of SUNNING emphasize much the importance of #12 Chalcedony. In addition, SUNNING accentuates the highly active positions of #1 Alunite (i.e., the scarlet dots), while the abundance maps of the competing methods in #3 Buddingtonite lack the precision as compared with SUNNING, such as SUBM–ADL and RHUIDR–HTV. On the other hand, the performance of MUA–BPT, CSUnL0–ADL, SUnSAL–ADL, LGSU, MdLRR, LSU, and MIHT drop significantly when Cuprite HSI is contaminated by sparse noise and stripe noise. Their corresponding abundance maps have almost provided no information.

To summarize, SUNNING achieves improved robust unmixing results in simulated noise cases compared with the state-of-the-art SU algorithms. More importantly, we favor SUNNING for SU problems because it does not involve sensitive and tedious tuning procedures with regularization parameters. On the contrary, JSTV, LRUnSAL-TV, SUBM-ADL, and

RHUIDR–HTV require adjustments to the regularization terms, in order to obtain their desirable robust effects. The proposed method is also different from RHUIDR–HTV and Pro–A–BM3D which require some strong prior assumptions and scenario information, such as including trimmed spectral library and noise level settings. While the settings for our SUNNING are pretty simple. A step in setting up sparsity level S is straightforward, as specified in Section III-B. For the robustness of hyperparameter a, we then discuss in Section V-F.

E. Effectiveness of SUNNING–SLP

Although SUNNING provides satisfying abundance maps in preserving the global structure without being affected by mixed noise, it sometimes loses the precision in fine details, as discussed in Section V-C. To deal with this, we conduct experiments on DC2 to investigate the robustness of SUNNING–SLP. Fig. 12 highlights that the abundance maps estimated by SUNNING–SLP are closer to the ground
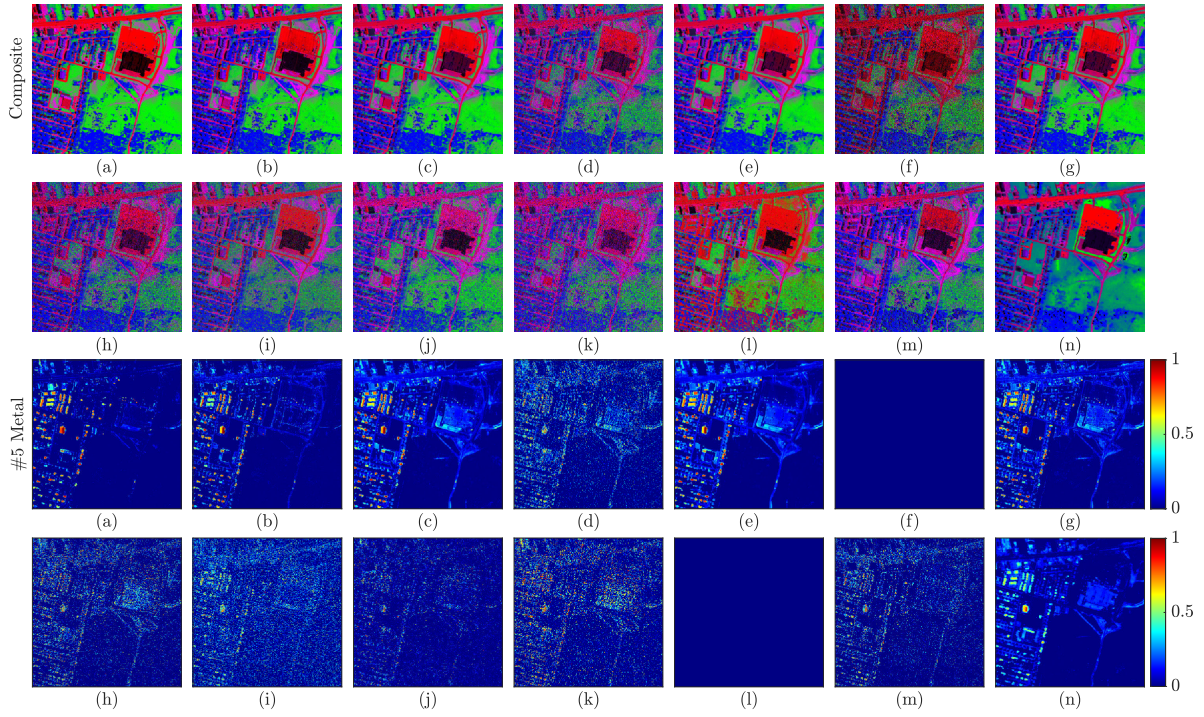
Fig. 8. Labeled reference and abundance maps obtained by various algorithms on Urban HSI under *Noise Case 1*. The first and second rows represent the composite abundance maps presented in pseudocolor, {#1 Red, #2 Green, #3 Blue, #4 Black, #5 Purple, #6 Magenta} correspond to the endmembers {#1 Asphalt, #2 Grass, #3 Tree, #4 Roof, #5 Metal, #6 Dirt}, respectively. The third and fourth rows represent the selected abundance maps of #5 Metal. (a) Reference. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.
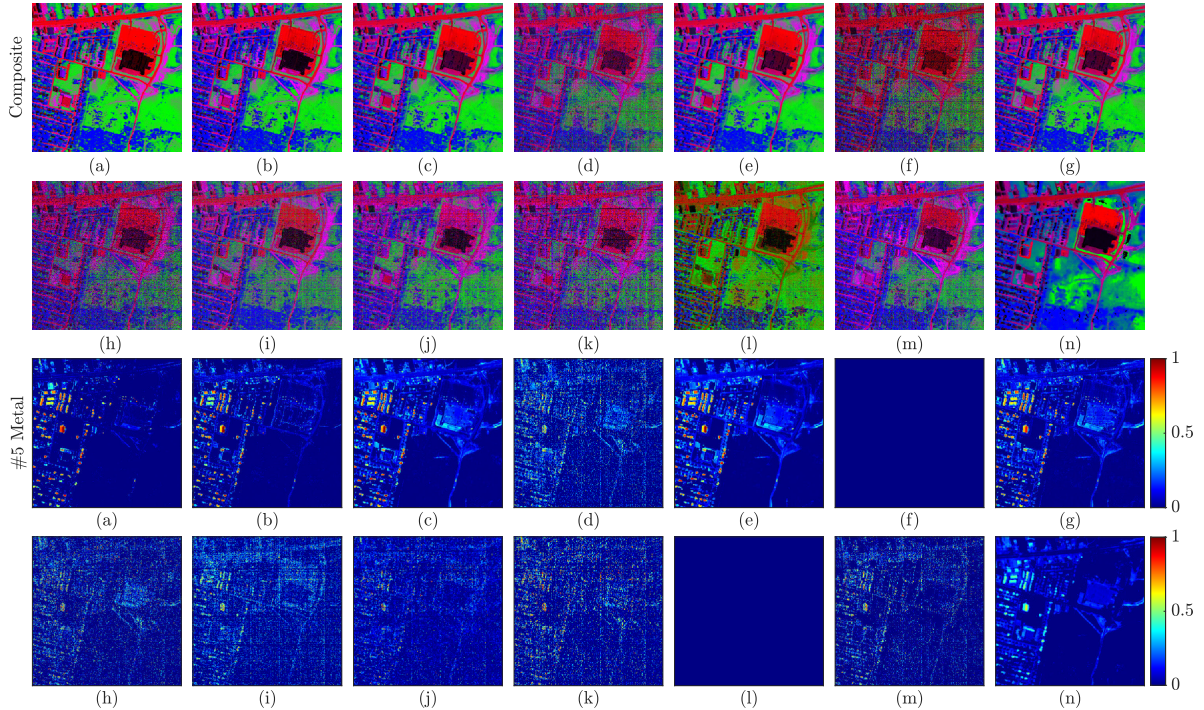

Fig. 9. Labeled reference and abundance maps obtained by various algorithms on urban HSI under *Noise Case 2*. (a) Reference. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.

truth. For example, in #1 Jarosite, we find that its abundance estimates are more accurate than those of SUNNING. Furthermore, the abundance variation and structural details of SUNNING–SLP are more enriched compared with the previous ones, and both active and inactive regions are improved. In the SLP strategy, a progressive dictionary update for ruling out confounders in $\boldsymbol{A}$ and $\boldsymbol{X}$ can boost the performance of SU. Enhanced SUNNING with the SLP strategy is demonstrated in Table V. Finally, we conduct experiments on noise-polluted Cuprite HSI and then evaluate
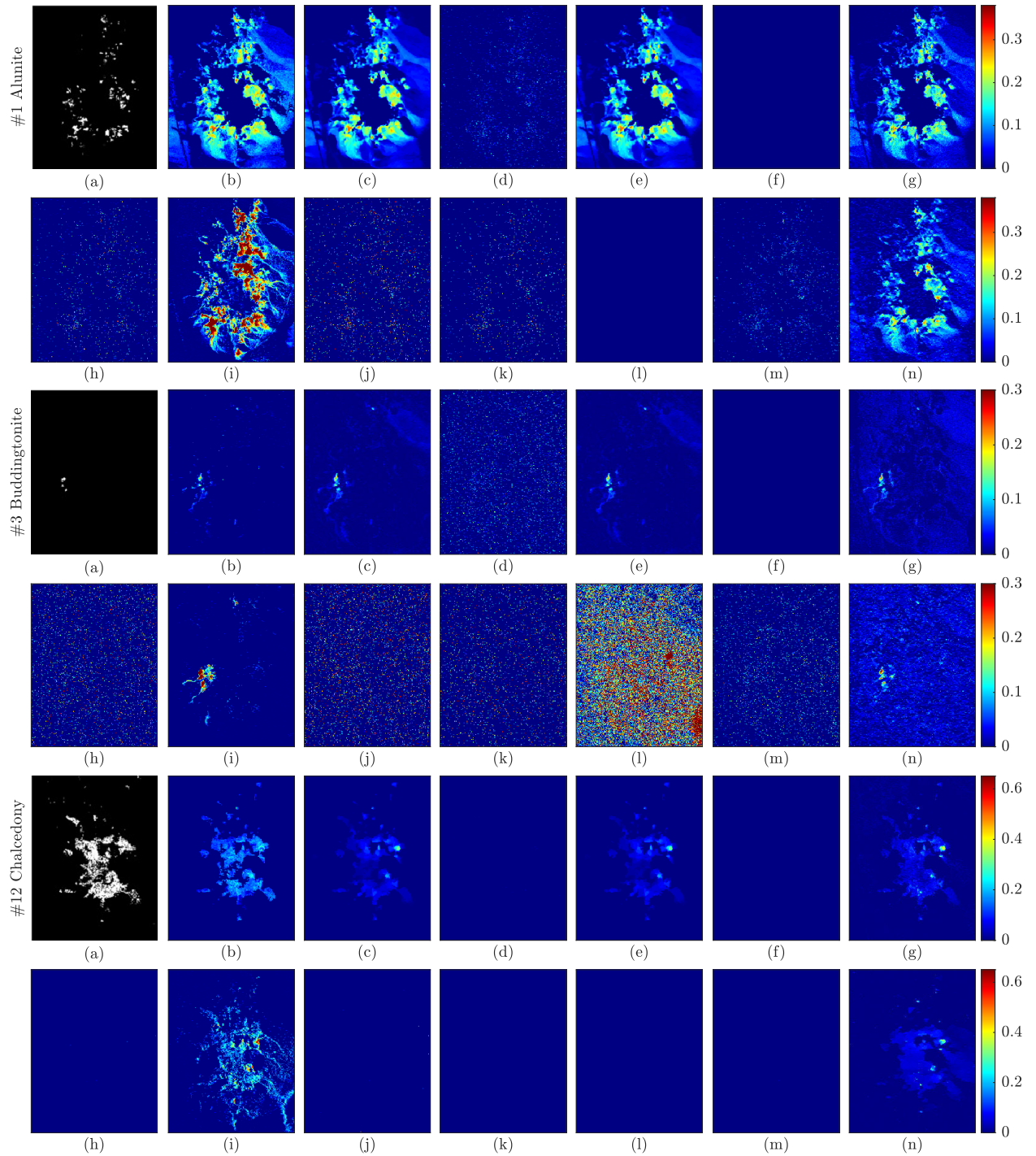
Fig. 10. Figure (a) in rows 1, 3, and 5 show the distribution maps of #1 Alunite, #3 Buddingtonite, and #12 Chalcedony. Figures (b)–(n) show Abundance maps obtained by various algorithms on Cuprite HSI under *Noise Case 1*. (a) Reference. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.

the performance of SUNNING–SLP. Owing to mixed noise, as shown in Fig. 10, the partial area is faint in the estimation of SUNNING, and the abundance variance is not obvious enough before the adoption of SLP. In *Noise Case 1*, Fig. 13 indicates that SUNNING–SLP can provide more abundance details than SUNNING when the irrelevant spectra are removed.

However, the assumption of SLP is not well considered in practice. As noticed in [12] and [22], there are calibration mismatches between the real spectra of Cuprite and the

spectral signatures available in the USGS library. SUNNING–SLP may prune endmember spectra due to the spectral signature mismatches. In real-world implementation, the mismatch problem may arise for some reasons [79], [80], such as the spectral similarities in the library, diverse physical conditions, calibration errors, and unpredictable factors. Due to these difficulties, SUNNING–SLP may fail to recover the true abundance values. The single SLP method may be unsuitable as the spectral signature mismatches influence abundance
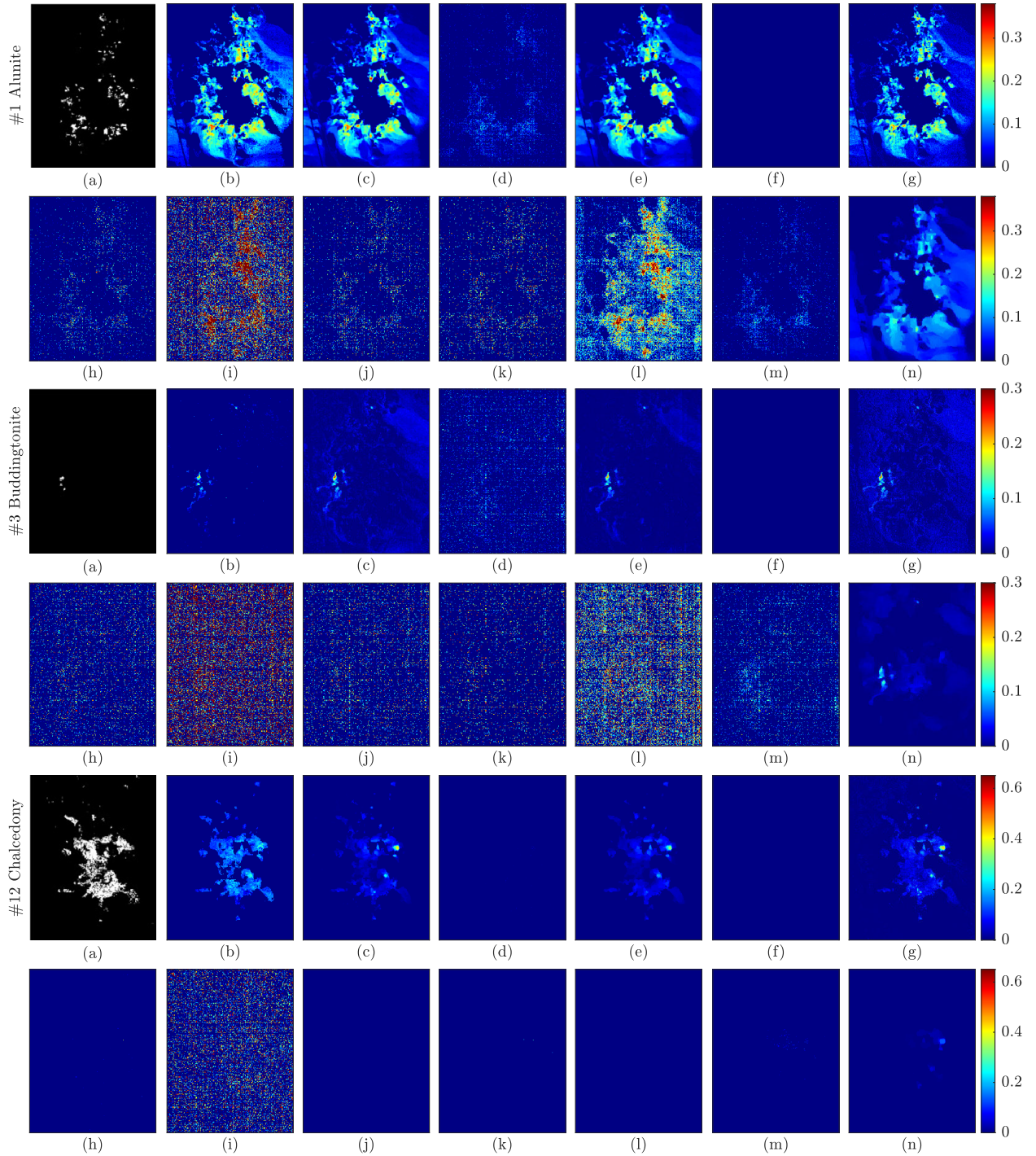
Fig. 11. Distribution maps and abundance maps estimated by various algorithms on Cuprite HSI under *Noise Case 2*. (a) Reference. (b) Our. (c) JSTV. (d) MUA–BPT. (e) LRUnSAL–TV. (f) CSUnL0–ADL. (g) SUBM–ADL. (h) SUnSAL–ADL. (i) Pro–A–BM3D. (j) LGSU. (k) MdLRR. (l) LSU. (m) MIHT. (n) RHUIDR–HTV.

estimation. To address such mismatch problems, the robust multiple signal classification (RMUSIC) [79] and the multiple dictionary pruning (MDP) [80] can be served as a powerful remedy in handling real-world cases.

F. Impact of Hyperparameter a

In this section, we investigate the impact of the hyperparameter a. As shown in Fig. 14, SRE is employed to evaluate the unmixing performance versus a in the presence of

mixed noise. It is seen that our SUNNING keeps improving SRE as the value of a increases. In particular, $a \rightarrow 100$ is sufficiently large to attain the optimal SRE. It is worth mentioning that SUNNING with $a = 1000$ is quite sensitive to small errors. As discussed in Section III-A, the gradient of $(1/a) \log(\cosh(a\epsilon))$ is $\tanh(a\epsilon)$. Considering $\tanh(100\epsilon)$ and $\tanh(1000\epsilon)$, the latter one produces a nearly constant, i.e., 1 or -1. Larger gradient values for small errors may affect the fitting process in gradient descent. As depicted in

TABLE IV
RUNTIME COMPARISON FOR VARIOUS SU
ALGORITHMS ON DC1 AND DC2

Algorithm	DC1		DC2	
	Iteration	Time(s)	Iteration	Time(s)
SUNNING	2000	119.04	2000	151.17
JSTV [26]	500	78.65	1000	228.52
MUA–BPT [61]	N/A	9.36	N/A	24.96
LRUnSAL–TV [62]	300	58.67	300	78.34
CSUnL0–ADL [39]	1500	119.92	1500	236.25
SUBM–ADL [39]	1000	457.11	1000	894.21
SUnSAL–ADL [39]	200	12.51	200	26.91
Pro–A–BM3D [63]	30	2.20	30	5.94
LGSU [64]	200	38.65	200	72.11
MdLRR [65]	300	51.33	300	85.70
LSU [66]	30	61.70	30	90.57
MIHT [57]	2000	134.54	2000	163.62
RHUIDR–HTV [67]	10000	1499.45	10000	3550.47

TABLE V
SREs, MSSIMs, AND THEIR STANDARD DERIVATIONS OF
SUNNING AND SUNNING–SLP ON DC2

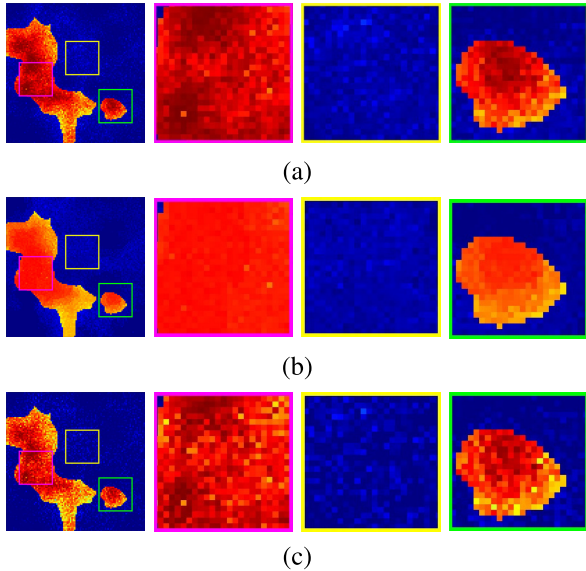| Noise	Metric	SUNNING	SUNNING–SLP
Noise Case 1	SRE	18.30±0.36	**20.90±0.64**
	MSSIM	0.58±0.01	**0.78±0.02**
Noise Case 2	SRE	18.33±0.39	**21.11±0.67**
	MSSIM	0.58±0.02	**0.79±0.02**

Fig. 12. Ground truth and estimated abundance maps correspond to #1 Jarosite and their cropped details on *Noise Case 2* of DC2. (a) Ground Truth and its cropped regions. (b) SUNNING and its cropped regions. (c) SUNNING–SLP and its cropped regions.

Fig. 14, we observe that there is a slight decrease in SRE when $a \to 1000$. To avoid this problem, we choose $a = 100$ in all experiments.
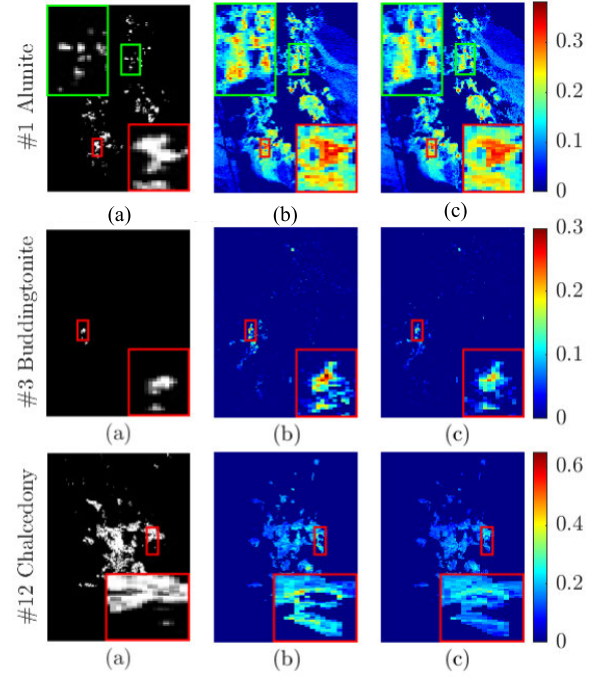

Fig. 13. Distribution maps and estimated abundance maps of SUNNING–SLP with $\phi = 0.01$ and $\Delta = 30$ and SUNNING on Cuprite HSI under *Noise Case 1*. (a) Reference. (b) SUNNING-SLP. (c) SUNNING.
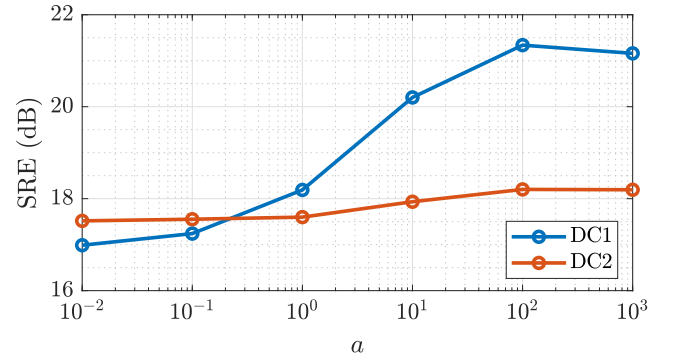

Fig. 14. SRE versus a under *Noise Case 1*.

Fig. 15. SRE versus Δ under *Noise Case 2*.
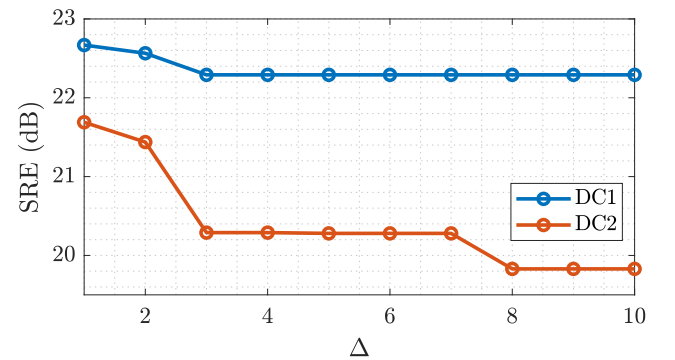
G. Parameter Setting of Δ

As recommended in [58], we empirically set $\phi = 0.02$ and $\Delta = 1$ in synthetic data experiments. Two experiments in *Noise Case 2* are conducted to evaluate the effect of parameter

Δ in SLP. In SUNNING–SLP, ϕ is an initial threshold. We initialize it to 0.02, in which any spectral signatures in $\boldsymbol{X}$ are distributed under 2% over the HSI, then it would be discarded from $\boldsymbol{A}$ and $\boldsymbol{X}$. The iterative threshold is updated by multiplying an iteration number q. In Fig. 15, as Δ increases, more spectral signatures are preserved in $\boldsymbol{A}$, which yields a slightly lower SRE. This behavior may be attributed to the relaxation of parameter S and the virtual abundance produced in estimation. Apparently, we find that SUNNING–SLP with $\Delta = 1$ or 10 provides a better reconstruction compared with SUNNING.

VI. CONCLUSION

In this article, we have developed SUNNING to address the hyperspectral unmixing problem in the presence of mixed noise. The proposed algorithm is derived from a PGD framework composed of gradient descent and nonconvex projection. There are two novelties in this work. First, SUNNING can explicitly control sparsity based on the need for a particular hyperspectral scene. The second is that we exploit a log-cosh function as objective loss, which is a generalization of the ℓ_2-norm and ℓ_1-norm objective loss functions. The hyperparameter a controls the effectiveness between ℓ_2-norm and ℓ_1-norm. In addition, we have provided the convergence proofs of SUNNING. Moreover, in the postprocessing phase, we consider the usefulness of the dictionary pruning strategy in SU. Adopting SUNNING with this technique can improve the accuracy and conditioning of SU. Experimental results on HSI datasets demonstrate that SUNNING and SUNNING–SLP both achieve more robust and reliable unmixing performance in the presence of different mixed noise scenarios.

APPENDIX A
PROOF OF THEOREM 1

Proof: Property 1) is straightforward to be verified. Let $\boldsymbol{d} \in \mathbb{R}^n$ be $\boldsymbol{Ax} - \boldsymbol{y}$. Given that $\log(\cosh(a\boldsymbol{d})) \geq \boldsymbol{0}$ for any $\boldsymbol{d} \in \mathbb{R}^n$, $\mathcal{G}(\boldsymbol{x}) = \sum_{i=1}^{L}(\log(\cosh(a(\boldsymbol{Ax} - \boldsymbol{y}))))/a \geq 0$ for any $\boldsymbol{x} \in \mathbb{R}^n$. Therefore, $\mathcal{G}(\boldsymbol{x})$ is lower bounded.

Next, we analyze Property 2) that the objective function $\mathcal{G}(\boldsymbol{x})$ is L-smooth. The Hessian of $\nabla^2\mathcal{G}(\boldsymbol{x})$ exists, that is, $\nabla^2\mathcal{G}(\boldsymbol{x}) = a\boldsymbol{A}^\top \text{diag}(\boldsymbol{1} - \tanh^2(a(\boldsymbol{Ax} - \boldsymbol{y})))\boldsymbol{A}$. It is clear that $\nabla^2\mathcal{G}(\boldsymbol{x}) \succeq \boldsymbol{0}$, where $\boldsymbol{1} - \tanh^2(a(\boldsymbol{Ax} - \boldsymbol{y})) \geq \boldsymbol{0}$. Herein, $\boldsymbol{0}$ denotes a zero matrix, and the symbol "$\succeq$" denotes ordering induced by positive semidefiniteness. For any two vectors $\boldsymbol{x}$, $\boldsymbol{y} \in \text{dom}\,\mathcal{G}$, let $\boldsymbol{z} = \boldsymbol{y} + h(\boldsymbol{x} - \boldsymbol{y})$, where $h \in [0, 1]$. We have

$$\nabla\mathcal{G}(\boldsymbol{x}) - \nabla\mathcal{G}(\boldsymbol{y}) = \nabla^2\mathcal{G}(\boldsymbol{z})(\boldsymbol{x} - \boldsymbol{y}). \quad (22)$$

Let $L > 0$ be the largest eigenvalue of $\nabla^2\mathcal{G}(\boldsymbol{x})$. Inducing the $\|\cdot\|_2$ norm to (22), such that

$$\|\nabla\mathcal{G}(\boldsymbol{x}) - \nabla\mathcal{G}(\boldsymbol{y})\|_2 = \left\|\nabla^2\mathcal{G}(\boldsymbol{z})(\boldsymbol{x} - \boldsymbol{y})\right\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2. \quad (23)$$

Based on (23), it is evident that $\mathcal{G}(\boldsymbol{x})$ is L-smooth. For simplicity, we consider $a\boldsymbol{A}^\top\boldsymbol{A}$ as the upper bound of $\nabla^2\mathcal{G}(\boldsymbol{x})$ because $a\boldsymbol{A}^\top\boldsymbol{A}$ does not depend on iteration number t and $\boldsymbol{x}^{(t)}$.

To prove Property 3), we consider the L-smoothness property of $\mathcal{G}(\boldsymbol{x})$. The following lemma is introduced.

Lemma 1: Given a L-smooth function $f : \mathbb{R}^n \to \mathbb{R}$, it has the quadratic upper bound property

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top(\boldsymbol{y} - \boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \quad (24)$$

where $\boldsymbol{x}, \boldsymbol{y} \in \text{dom}(f)$ and $L > 0$ is a Lipschitz constant.

From Lemma 1, we have

$$\begin{aligned}
&\mathcal{G}(\boldsymbol{x}^{(t+1)}) - \mathcal{G}(\boldsymbol{x}^{(t)}) \\
&\leq \nabla\mathcal{G}(\boldsymbol{x}^{(t)})^\top(\boldsymbol{x}^{(t+1)} - \boldsymbol{z}^{(t+1)}) - \eta\|\nabla\mathcal{G}(\boldsymbol{x}^{(t)})\|_2^2 \\
&\quad + \frac{L}{2}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \\
&= \frac{1}{2\eta}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{z}^{(t+1)}\|_2^2 + \frac{1}{2\eta}\|\boldsymbol{z}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \\
&\quad - \frac{1}{2\eta}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 - \eta\|\nabla\mathcal{G}(\boldsymbol{x}^{(t)})\|_2^2 \\
&\quad + \frac{L}{2}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \\
&\leq \frac{1}{\eta}\|\boldsymbol{z}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 - \eta\|\nabla\mathcal{G}(\boldsymbol{x}^{(t)})\|_2^2 \\
&\quad + \left(\frac{L}{2} - \frac{1}{2\eta}\right)\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \\
&= \left(\frac{L}{2} - \frac{1}{2\eta}\right)\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \quad (25)
\end{aligned}$$

where $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{z}^{(t+1)}\|_2^2 \leq \|\boldsymbol{z}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2$ implicitly holds since $\boldsymbol{x}^{(t+1)}$ is an optimal solution to $\mathcal{P}_C(\boldsymbol{z}^{(t+1)})$.

According to (25), we further express the inequality with $\eta \in (0, 1/L)$ as follows:

$$\mathcal{G}(\boldsymbol{x}^{(t+1)}) - \mathcal{G}(\boldsymbol{x}^{(t)}) \leq 0. \quad (26)$$

We conclude that if $0 < \eta < 1/L$ and $\boldsymbol{x}^{(t+1)} \neq \boldsymbol{x}^{(t)}$, then $\mathcal{G}(\boldsymbol{x}^{(t+1)}) < \mathcal{G}(\boldsymbol{x}^{(t)})$. The proof is complete. ∎

APPENDIX B
PROOF OF THEOREM 2

Proof: Given that $\|\boldsymbol{x}^{(0)}\|_2^2 < +\infty$, if $\|\boldsymbol{x}^{(t+1)}\|_2 \to +\infty$, then $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \to +\infty$ must hold. In addition, $\mathcal{G}(\boldsymbol{x}^{(t)}) \to +\infty$ holds, since $\mathcal{G}(\boldsymbol{x}^{(t)}) \geq \mathcal{G}(\boldsymbol{x}^{(t+1)}) + (1/(2\eta) - L/2)\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2$ from (25), contradicting Theorem 1. Thereby, the objective function $\mathcal{G}(\boldsymbol{x})$ is radial unbounded.

For Property 2), as $\boldsymbol{x}^{(t+1)} \in C$, i.e., $\boldsymbol{1}^\top\boldsymbol{x}^{(t+1)} = 1$, and $\boldsymbol{x}^{(t+1)} \geq 0$, hence $\boldsymbol{x}^{(t+1)}$ is bounded.

For Property 3), based on Theorem 1, we prove that $\{\mathcal{G}(\boldsymbol{x}^{(t)})\}_{t=0}^{\infty}$ is lower bounded, and its sequence is convergent which asymptotically leads to $\boldsymbol{x}^*$, such that $\mathcal{G}^* = \inf\mathcal{G}(\boldsymbol{x})$ is finite. Summing over t from 0 to $N - 1$, (25) becomes

$$\sum_{t=0}^{N-1} c\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \leq \sum_{t=0}^{N-1}[\mathcal{G}(\boldsymbol{x}^{(t)}) - \mathcal{G}(\boldsymbol{x}^{(t+1)})]$$

$$\sum_{t=0}^{N-1}\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2 \leq \frac{1}{c}(\mathcal{G}(\boldsymbol{x}^{(0)}) - \mathcal{G}(\boldsymbol{x}^{(N)})) \quad (27)$$

where c denotes a positive constant of $1/(2\eta) - L/2$.

Then, consider the limit of (27) when N goes to infinity and $\mathcal{G}(\boldsymbol{x}^{(N)})$ reaches the minimum $\mathcal{G}^*$. We have

$$\lim_{N \to \infty} \sum_{t=0}^{N} \left\| \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \right\|_2^2 \leq \frac{1}{c} \left(\mathcal{G}(\boldsymbol{x}^{(0)}) - \mathcal{G}^* \right) < +\infty \quad (28)$$

$$\Rightarrow \lim_{t \to \infty} \left\| \boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)} \right\|_2^2 = 0. \quad (29)$$

In (28), it shows that the infinite sum of $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\|_2^2$ is bounded below $+\infty$. It implies (29) that when $N \to \infty$ and $\mathcal{G}(\boldsymbol{x}^{(t+1)}) - \mathcal{G}(\boldsymbol{x}^{(t)}) \leq 0$ holds, the residual between $\boldsymbol{x}^{(t+1)}$ and $\boldsymbol{x}^{(t)}$ will tend toward 0. Otherwise, it will contradict with the convergent series of $\{\mathcal{G}(\boldsymbol{x}^{(t)})\}_{t=0}^{\infty}$.

Due to (29) and the boundedness of $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$, given $\boldsymbol{x}^{(0)}$, $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$ has the convergent subsequences, and each of subsequences has its limit. Based on the convergent behavior stated in Theorem 1, thereby all the subsequences would end up with the same limit. Otherwise, $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$ with different limits will alternatively change the objective function value $\mathcal{G}(\boldsymbol{x}^{(t)})$ and violate Theorem 1. As a result, we prove that $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$ has one limit only. Proof of Theorem 2 is complete. ∎

REFERENCES

[1] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, Aug. 2020.

[2] Y. Yuan, C. Wang, and Z. Jiang, "Proxy-based deep learning framework for spectral–spatial hyperspectral image classification: Efficient and robust," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501115.

[3] W. Cao, K. Wang, G. Han, J. Yao, and A. Cichocki, "A robust PCA approach with noise structure learning and spatial–spectral low-rank modeling for hyperspectral image restoration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3863–3879, Oct. 2018.

[4] Y. Chen, W. Cao, L. Pang, J. Peng, and X. Cao, "Hyperspectral image denoising via texture-preserved total variation regularizer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5516114.

[5] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.

[6] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.

[7] W. Dong et al., "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[8] H.-X. Dou, Z.-C. Wu, Y.-W. Zhuo, L.-J. Deng, and G. Vivone, "Dynamical fusion model with joint variational and deep priors for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[9] Z. Yang, M. Xu, S. Liu, H. Sheng, and H. Zheng, "Spatial–spectral attention bilateral network for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[10] T. Ince and N. Dobigeon, "A fast spatial–spectral NMF for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[11] D. C. Heinz and Chein-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.

[12] 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.

[13] 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.

[14] 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.

[15] S. Zhang, A. Agathos, and J. Li, "Robust minimum volume simplex analysis for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6431–6439, Nov. 2017.

[16] 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.

[17] X. Li, X. Zhang, Y. Yuan, and Y. Dong, "Adaptive relationship preserving sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5504516.

[18] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.

[19] T. Ince and T. Dundar, "Simultaneous nonconvex denoising and unmixing for hyperspectral imaging," *IEEE Access*, vol. 7, pp. 124426–124440, 2019.

[20] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral subspace identification using SURE," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2481–2485, Dec. 2015.

[21] C. Y. Zheng, H. Li, Q. Wang, and C. L. Philip Chen, "Reweighted sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 479–488, Jan. 2016.

[22] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[23] 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.

[24] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral a priori information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 770–783, Feb. 2015.

[25] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, Sep. 2009.

[26] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4257–4266, Sep. 2016.

[27] L. Zhuang and M. K. Ng, "Hyperspectral mixed noise removal by ℓ_1-norm-based subspace representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1143–1157, 2020.

[28] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1128–1141, Sep. 2015.

[29] X. P. Li, Z.-L. Shi, C.-S. Leung, and H. C. So, "Sparse index tracking with K-sparsity or e-Deviation constraint via l0-norm minimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 1–14, Dec. 2022.

[30] 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.

[31] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using l0 norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5495–5508, Sep. 2018.

[32] J. M. Bioucas-Dias and M. A. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Jun. 2010, pp. 1–4.

[33] 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.

[34] S. Wang, T.-Z. Huang, X.-L. Zhao, G. Liu, and Y. Cheng, "Double reweighted sparse regression and graph regularization for hyperspectral unmixing," *Remote Sens.*, vol. 10, no. 7, p. 1046, Jul. 2018.

[35] L. Ren, D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot, "Hyperspectral sparse unmixing via nonconvex shrinkage penalties," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500415.

[36] C. Li, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Sparse unmixing of hyperspectral data with noise level estimation," *Remote Sens.*, vol. 9, no. 11, p. 1166, Nov. 2017.

[37] C. Li et al., "Sparse unmixing of hyperspectral data with bandwise model," *Inf. Sci.*, vol. 512, pp. 1424–1441, Feb. 2020.

[38] Y. Ma, C. Li, X. Mei, C. Liu, and J. Ma, "Robust sparse hyperspectral unmixing with $\ell_{2,1}$ norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, Dec. 2016.

[39] X. Zhang, Y. Yuan, and X. Li, "Sparse unmixing based on adaptive loss minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5531514.

[40] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[41] B. Chen, L. Xing, H. Zhao, S. Du, and J. C. Príncipe, "Effects of outliers on the maximum correntropy estimation: A robustness analysis," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 6, pp. 4007–4012, Jun. 2021.

[42] F. Zhu, A. Halimi, P. Honeine, B. Chen, and N. Zheng, "Correntropy maximization via ADMM: Application to robust hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4944–4955, Sep. 2017.

[43] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Ann. Data Sci.*, vol. 9, no. 2, pp. 187–212, Apr. 2020.

[44] O. Karal, "Maximum likelihood optimal and robust support vector regression with lncosh loss function," *Neural Netw.*, vol. 94, pp. 1–12, Oct. 2017.

[45] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7.

[46] L. Yu, L. Xu, and X. Jiang, "An effective method for detecting unknown types of attacks based on log-cosh variational autoencoder," *Appl. Sci.*, vol. 13, no. 22, p. 12492, Nov. 2023.

[47] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, Nov. 2018.

[48] R. A. Saleh and A. K. M. E. Saleh, "Statistical properties of the log-cosh loss function used in machine learning," 2022, *arXiv:2208.04564*.

[49] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2001.

[50] K. Gokcesu and H. Gokcesu, "Generalized Huber loss for robust learning and its efficient minimization for a robust statistics," 2021, *arXiv:2108.12627*.

[51] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[52] 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.

[53] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 629–654, Dec. 2008.

[54] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[55] J. Chen, S. Yang, Z. Wang, and H. Mao, "Efficient sparse representation for learning with high-dimensional data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 1–15, Oct. 2021.

[56] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Registered Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[57] X.-P. Li, C.-S. Leung, and H. C. So, "Sparse recovery under nonnegativity and sum-to-one constraints," *Inf. Sci.*, vol. 679, Sep. 2024, Art. no. 121059.

[58] X. Zhang et al., "Hyperspectral unmixing via low-rank representation with space consistency constraint and spectral library pruning," *Remote Sens.*, vol. 10, no. 2, p. 339, Feb. 2018.

[59] X. Xu, B. Pan, Z. Chen, Z. Shi, and T. Li, "Simultaneously multiobjective sparse unmixing and library pruning for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3383–3395, Apr. 2021.

[60] X. Zhang, Y. Yuan, and X. Li, "Reweighted low-rank and joint-sparse unmixing with library pruning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527816.

[61] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, and C. Richard, "A fast multiscale spatial regularization for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 598–602, Apr. 2019.

[62] T. Ince, "Robust hyperspectral unmixing using total variation regularized low-rank approximation," in *Proc. 9th Int. Conf. Recent Adv. Space Technol. (RAST)*, Jun. 2019, pp. 373–379.

[63] M. Zhao, X. Wang, J. Chen, and W. Chen, "A plug-and-play priors framework for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501213.

[64] X. Shen, H. Liu, X. Zhang, K. Qin, and X. Zhou, "Superpixel-guided local sparsity prior for hyperspectral sparse regression unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[65] L. Wu, J. Huang, and M.-S. Guo, "Multidimensional low-rank representation for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[66] X. Shen et al., "Efficient hyperspectral sparse regression unmixing with multilayers," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522614.

[67] K. Naganuma and S. Ono, "Toward robust hyperspectral unmixing: Mixed noise modeling and image-domain regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 8117–8138, 2024.

[68] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[69] 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.

[70] 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.

[71] 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.

[72] F. Zhu, "Hyperspectral unmixing: Ground truth labeling, datasets, benchmark performances and survey," 2017, *arXiv:1708.05125*.

[73] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.

[74] X.-R. Feng, H.-C. Li, R. Wang, Q. Du, X. Jia, and A. Plaza, "Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4414–4436, 2022.

[75] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.

[76] Z.-L. Shi, X. P. Li, C.-S. Leung, and H. C. So, "Cardinality constrained portfolio optimization via alternating direction method of multipliers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1–9, Feb. 2022.

[77] R. N. Clark et al., "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *J. Geophys. Res., Planets*, vol. 108, no. E12, pp. 1–24, Dec. 2003.

[78] Y. Zheng, F. Wu, H. J. Shim, and L. Sun, "Sparse unmixing for hyperspectral image with nonlocal low-rank prior," *Remote Sens.*, vol. 11, no. 24, p. 2897, Dec. 2019.

[79] X. Fu, W.-K. Ma, J. M. Bioucas-Dias, and T.-H. Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5171–5184, Sep. 2016.

[80] P. Wang, X. Shen, K. Ni, and L. Shi, "Hyperspectral sparse unmixing based on multiple dictionary pruning," *Int. J. Remote Sens.*, vol. 43, no. 7, pp. 2712–2734, Apr. 2022.

Yiu Yu Chan received the B.Eng. degree in computer and data engineering from the City University of Hong Kong, Hong Kong, SAR, China, in 2022, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, supervised by Prof. Chi-Sing Leung.

His research interests include neural networks, machine learning, and optimization.

Xiao-Peng Li (Member, IEEE) received the B.Eng. degree (Hons.) in electronic science and technology from Yanshan University, Qinhuangdao, China, in 2015, and the M.Sc. degree (Hons.) in electronic information engineering and the Ph.D. degree in electrical engineering from the City University of Hong Kong, Hong Kong, SAR, China, in 2018 and 2022, respectively.

He was a Research Assistant with the Department of Information Engineering, Shenzhen University, Shenzhen, China, from 2018 to 2019, and a Post-Doctoral Fellow with the Department of Electrical Engineering, City University of Hong Kong, from 2022 to 2023. He is currently an Assistant Professor with the College of Electronics and Information Engineering, Shenzhen University. His research interests include robust signal processing, sparse recovery, matrix processing, tensor processing, optimization methods, machine learning, and their applications in various areas of engineering, including target estimation, image recovery, video restoration, hyperspectral unmixing, and stock market analysis.

Jiajie Mai received the B.E. degree in telecommunications engineering with management from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2020, and the M.Sc. degree in mobile and personal communication from the King's College London, London, U.K., in 2021. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, SAR, China, supervised by Prof. Chi-Sing Leung.

His research interests include neural networks, machine learning, and optimization.

Chi-Sing Leung (Senior Member, IEEE) received the Ph.D. degree in computer science from The Chinese University of Hong Kong, Hong Kong, in 1995.

He is currently a Professor with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. He has authored or co-authored more than 120 journal articles in the areas of digital signal processing, neural networks, and computer graphics. His research interests include neural computing and computer graphics.

Dr. Leung was a recipient of the 2005 IEEE Transactions on Multimedia Prize Paper Award for his paper titled "The Plenoptic Illumination Function." He was an Organizing Committee Member of ICONIP 2006. He was the Program Chair of ICONIP 2009 and ICONIP 2012. He is a Governing Board Member and a Vice President of the Asian–Pacific Neural Network Assembly (APNNA). He is/was a Guest Editor for several journals, including *Neural Computing and Applications*, *Neurocomputing*, and *Neural Processing Letters*.
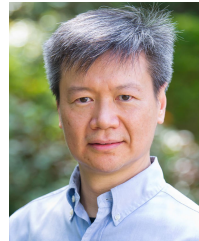
Hing Cheung So (Fellow, IEEE) was born in Hong Kong. He received the B.Eng. degree in electronic engineering from the City University of Hong Kong and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Hong Kong, in 1990 and 1995, respectively.

From 1990 to 1991, he was an Electronic Engineer with the Research and Development Division, Everex Systems Engineering Ltd., Hong Kong. From 1995 to 1996, he was a Post-Doctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor. His research interests include detection and estimation, fast and adaptive algorithms, multidimensional harmonic retrieval, robust signal processing, source localization, and sparse approximation.

Dr. So was on the editorial boards of *IEEE Signal Processing Magazine* from 2014 to 2017, IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2010 to 2014, *Signal Processing* since 2010, and *Digital Signal Processing* since 2011. He was also a Lead Guest Editor for a Special Issue on "Advances in Time/Frequency Modulated Array Signal Processing" of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, in 2017. In addition, he was an Elected Member of Signal Processing Theory and Methods Technical Committee from 2011 to 2016 of the IEEE Signal Processing Society where he was the Chair in the awards subcommittee from 2015 to 2016.