Unsupervised Feature Dimensionality Reduction via Latent Low-Rank Embedding Projection for Classification of Hyperspectral Images

Heng-Chao Li, Jun-Qiu Wang, Si-Jia Xiang, Jing-Hua Yang, and Qian Du

***Abstract*—To deal with the curse of dimensionality in hyperspectral images, numerous feature dimensionality reduction (FDR) methods have been proposed to map high-dimensional data into a low-dimensional subspace. However, most of existing FDR methods lack robustness against noise corruption. To this end, the representation-based subspace learning has been developed to find a robust projection matrix for FDR. Nevertheless, most of them only consider a single direction of the matrix, which ignore the information from other directions. Moreover, the majority of existing methods fail to account for both global structure and feature correlations effectively. To address the above problems, we propose a novel robust projection learning method called latent low-rank embedding (LatLRE), which integrates the latent low-rank representation (LatLRR) with projection learning. In particular, the proposed model can maintain the strong robustness of LatLRR and simultaneously learn a projection for FDR. Moreover, the nuclear norm and logarithmic norm are employed to approximate the two underlying rank functions and provide a more accurate measure of correlation. In addition, LatLRE is optimized using the alternating direction method of multipliers (ADMM) algorithm with the theoretical convergence guarantee. To verify the FDR performance of LatLRE, extensive experiments are conducted on three benchmark hyperspectral datasets. The experimental results demonstrate that LatLRE outperforms other FDR methods considered in this paper.**

***Index Terms*—Hyperspectral images (HSIs), feature dimensionality reduction, pattern recognition, robust projection learning (RPL), latent low-rank representation.**

I. Introduction

THE advent of hyperspectral sensors has enabled the acquisition of hyperspectral images (HSIs) that capture comprehensive spectral and spatial information simultaneously [1], [2], [3], [4], [5], [6], [7], [8]. This detailed spectral information is captured across hundreds of bands, which has found extensive use in diverse fields, including forestry studies, precision agriculture, and environmental monitoring [9], [10]. Despite the richness of high-dimensional hyperspectral information, it also engenders the curse of dimensionality, which presents a formidable challenge to the generalization capability of model. Additionally, the high dimensionality of HSIs directly leads to a dramatic increase in computational complexity and memory consumption. Besides, the redundancy contained in high-dimensional features usually degrades the performance of postprocessing algorithms [11].

Over the past decades, significant progress has been made in the development of feature dimensionality reduction (FDR) technology [12], [13], [14], [15], [16], [17], [18]. The typical FDR methods can be grouped into three categories: supervised, semi-supervised, and unsupervised strategies. Although the supervised and semi-supervised FDR algorithms have demonstrated outstanding performance in the literatures [14], [15], [16], they depend on the data labeling process and are often required to solve difficult non-convex and non-smooth optimization problems. Therefore, unsupervised FDR methods have received considerable attention due to their ease of implementation. The most typical unsupervised FDR method is principal component analysis (PCA) [17], which attempts to find an optimal projection matrix such that the projected data have the maximum variance. To improve the FDR performance, some robust distances were introduced into PCA for model development [18], [19], [20], [21], [22]. Recently, several enhanced PCA variants have been proposed and have shown good performance in HSIs. One such variant is SuperPCA [23], which partitions the HSIs into homogeneous regions and applies PCA in each superpixel block to extract local features. Nevertheless, this approach ignores the global information. To address this limitation, S^3 PCA [24] has been developed, which incorporates a complement to the global features. However, above PCA and its extensions ignore the nonlinear structure information in data.

Since nonlinear techniques have the merit of preserving geometrical structure of data manifold, the corresponding algorithms, such as locality preserving projections (LPP) [25], [26], [27] and preserving neighborhood discriminant embedding (PNDE) [28], have been widely used for feature extraction (FE) in HSIs. Nonetheless, the selection of neighborhood size of LPP is a challenge. Zhang *et al.* [29] integrated nuclear norm and $L_{2,1}$-norm constraints into 2D neighborhood preserving projections (2DNPP) [30], and developed Nuclear norm-based 2DNPP. In [31], Yan *et al.* claimed that the previous methods, such as PCA and LPP can be unified into a general dimensionality reduction framework termed graph embedding (GE). The GE-based models [32], [33] have been widely employed for the efficient representation of local manifold structures. Although the GE-based methods show

This work was supported by the National Natural Science Foundation of China under Grants 62271418 and 12401605, the Natural Science Foundation of Sichuan Province under Grants 2023NSFSC0030 and 2024NSFSC1467, Postdoctoral Fellowship Program of CPSF under Grant GZC20232198, China Postdoctoral Science Foundation under Grant 2024M752661, and State Key Laboratory of Intelligent Geotechnics and Tunnelling (FSDI). (Corresponding author: Jing-Hua Yang)

Heng-Chao Li, Jun-Qiu Wang, Si-Jia Xiang, Jing-Hua Yang are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China, (e-mail: lihengchao_78@163.com and yangjinghua110@126.com).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762 USA.

promising performance, the intrinsic graph is sensitive to various kinds of noise, such as corruptions and occlusions.

Recently, the sparse representation (SR) technique has drawn great attention in image processing. SR focuses on representing a signal or data point using a minimal number of basis elements or features from a dictionary. In detail, only a few components of the coefficients are non-zero, and the rest are zeros. From the view of classification, Refs. [34] and [35] have demonstrated that SR is advantageous to achieve high accuracy in signal classification and face recognition, even under high-level noise corruptions and occlusions. Therefore, a series of FDR methods have been proposed based on SR [36], [37]. Undoubtedly, these methods have further promoted the development of the FDR techniques. However, SR-based FDR cannot reveal the global structure of data and may degrade the performance when the training data are corrupted by complex and correlated noise.

To overcome the shortcoming of the sparse-based FDR and exploit the global structure of the data, Liu *et al.* [38], [39] proposed a robust method named low-rank representation (LRR). LRR aims at uncovering the lowest-rank subspace of data and removing the adverse impacts of the noise and outliers in samples. Under the assumption that samples from the same category lie on an inherent low-dimensional subspace, LRR and its variants [40], [41] can maintain the underlying global structure, *i.e.*, low-dimensional multi-subspace structure. In order to address challenges posed by high-dimensional problems, the methods of LPP have been combined with projection learning to enhance robustness, such as LRE [42] and LSPP [43]. Subsequently, to address the case of insufficient sampling or grossly corrupted data, Liu *et al.* [44] proposed latent low-rank representation (LatLRR), which is developed on the presupposition that unobserved samples can be used to represent the observed samples. Furthermore, numerous methods based on LatLRR [45], [46], [47] have been developed and implemented with promising performance. Nevertheless, the aforementioned methods overlook the issues of robust feature extraction problem, and may not fully explore the correlations between features.

To address high dimensionality problem and improve noise robustness, we propose a novel latent low-rank embedding (LatLRE) for robust FDR by combining LRR with projection learning, see Fig. 1. As discussed before, LatLRR is robust to the corrupted and hidden data. Correspondingly, the proposed LatLRE integrates LatLRR with projection learning together into one model, which retains the robustness of LatLRR to learn a low-dimensional projection matrix for enhancing HSIs classification. In summary, the main contributions of this paper are summarized as follows:

- We propose a novel Latent Low-Rank Embedding (LatLRE) method for unsupervised robust feature dimensionality reduction (FDR) in HSI classification. Our LatLRE leverages the latent low-rank constraint to reveal the potential data structure and employs the robust $\ell_{2,1}$-norm to measure the reconstruction error. Furthermore, by integrating projection learning with LatLRR, the proposed LatLRE can explore both row and column information simultaneously, ensuring robustness in the low-dimensional feature space.
- Considering that rank minimization is a nonconvex optimization problem, we employ nuclear norm and logarithmic norm in LatLRE to approximate the two underlying rank functions. The adoption of dual-norm approximation not only reveals the global data structure more clearly during reconstruction but also reduces redundant information and better exploits feature correlations.
- An iterative algorithm based on the alternating direction method of multipliers (ADMM) framework is designed to solve the LatLRE model. Moreover, for the proposed nonconvex model, we prove the boundedness of the generated sequence and the convergence of the iterative algorithm. In addition, we provide a computational complexity analysis. Extensive experiments on several benchmark HSIs datasets are conducted to validate the effectiveness and robustness of the proposed method in classification tasks.

The remainder of this paper is organized as follows. Section II introduces the related works. Then, Section III elaborates the proposed model and its optimization. The experiments and analysis are presented in Section IV. Finally, Section V gives the conclusion of this paper.

II. Related Works

A. Low-Rank Embedding

The LRR [38] can reveal the potential low-rank representation of data, but the dimensionality of the obtained representation vectors is highly related to the size of the dictionary. To this end, LRE [42] was proposed by integrating LRR and projection learning together to address the robust feature dimensionality reduction problem. LRE can overcome the "curse of dimensionality" problem through the learned low-dimensional projection. The objective function of LRE is formulated as

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{P}} \ rank(\mathbf{Z}) + \lambda \left\| \mathbf{P}^T\mathbf{X} - \mathbf{P}^T\mathbf{X}\mathbf{Z} \right\|_{2,1}, \\ &s.t.\ \mathbf{P}^T\mathbf{P} = \mathbf{I}, \end{aligned} \tag{1}$$

where $rank(\mathbf{Z})$ denotes the rank of matrix $\mathbf{Z}$, $\mathbf{P}$ is the low-dimensional projection matrix, $\mathbf{I}$ is an identity matrix, and $\|\cdot\|_{2,1}$ represents $\ell_{2,1}$-norm. For avoiding the NP-hard problem, the $rank(\cdot)$ in Eq. (1) is substituted with nuclear norm. To enhance the robustness on the noise and outliers, $\ell_{2,1}$-norm is employed to measure the reconstruction error rather than ℓ_1-norm, which is defined as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{m} \mathbf{X}_{ij}^2}$. Therefore, the following formulation can be obtained:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{P}} \ \|\mathbf{Z}\|_* + \lambda \left\| \mathbf{P}^T\mathbf{X} - \mathbf{P}^T\mathbf{X}\mathbf{Z} \right\|_{2,1}, \\ &s.t.\ \mathbf{P}^T\mathbf{P} = \mathbf{I}. \end{aligned} \tag{2}$$

For ease of mathematical derivation, a variable $\mathbf{E} = \mathbf{P}^T\mathbf{X} - \mathbf{P}^T\mathbf{X}\mathbf{Z}$ is introduced, which is utilized to measure the reconstructive property and other noise. Thus, the objective function (2) can be rewritten as

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{P}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \\ &s.t.\ \mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{X}\mathbf{Z} - \mathbf{E},\ \mathbf{P}^T\mathbf{P} = \mathbf{I}. \end{aligned} \tag{3}$$

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2025.3627780

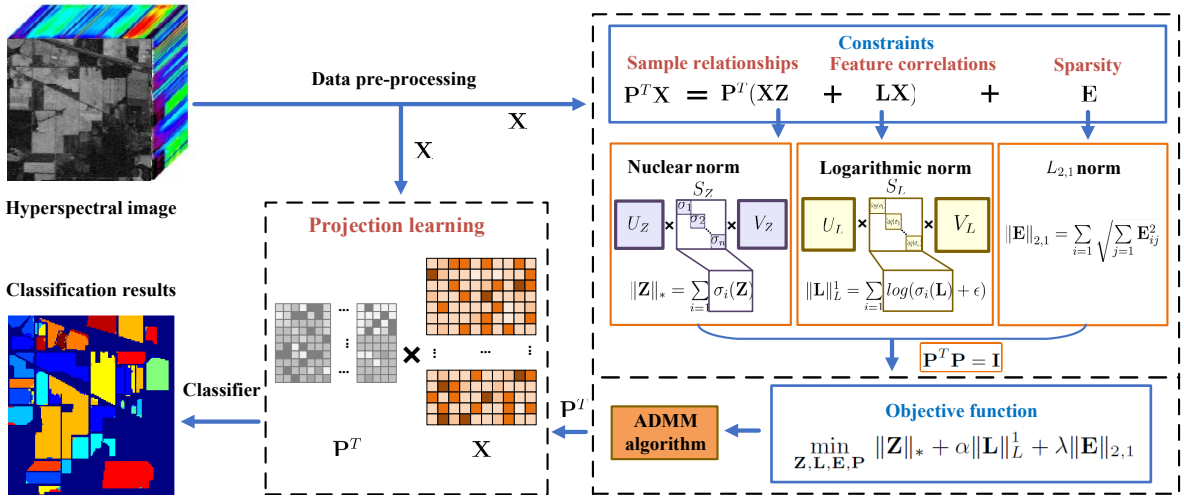


Fig. 1. The framework of LatLRE. First, LatLRE decomposes the hyperspectral image data into principal feature representation $\mathbf{XZ}$, salient feature representation $\mathbf{LX}$, as well as noise matrix $\mathbf{E}$. Then, we introduce a projection matrix $\mathbf{P}$ to achieve feature dimensionality reduction. To depict the low-rankness of the sample relationships $\mathbf{Z}$ and feature correlations $\mathbf{L}$, we impose a nuclear norm on $\mathbf{Z}$, and a logarithmic norm on $\mathbf{L}$. Moreover, $l_{2,1}$ norm is introduced on noise matrix $\mathbf{E}$ to improve the robustness. Finally, the classification results can be obtained by support vector machine (SVM) classifier.

Then, the above model can be solved by an alternative iteration algorithm [44].

B. Latent Low-Rank Representation

The goal of LRR [38] is to learn a low-rank coefficient matrix for capturing the underlying structure of the given data. In particular, the objective function of LRR is formulated as

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1, \ s.t. \ \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{4}$$

where $\mathbf{X} \in \mathbb{R}^{m\times N}$ denotes input data matrix, $\mathbf{Z} \in \mathbb{R}^{N\times N}$ is the reconstruction matrix, λ is the regularization parameter, m is the dimensionality of each data point, and N is the number of data points. $\|\mathbf{Z}\|_*$ is the nuclear norm of $\mathbf{Z}$, which is computed by $\|\mathbf{Z}\|_* = \sum_i \sigma_i$, where σ_i is the ith singular value of $\mathbf{Z}$. $\mathbf{E}$ represents the noisy data matrix. $\|\mathbf{E}\|_1$ denotes the ℓ_1-norm of $\mathbf{E}$, whose definition is $\|\mathbf{E}\|_1 = \max_{1\le j\le N} \sum_{i=1}^{N} |\mathbf{E}_{ij}|$. Generally, the explicit data itself is selected as the dictionary for LRR. Refs. [38] and [39] have shown that LRR is an effective method for both subspace clustering and data reconstruction. However, the performance of LRR is severely affected when the data is corrupted by the high level of noise, such as block occlusions. To alleviate the impact of noise, Liu *et al.* [44] proposed LatLRR by constructing the dictionary with both explicit and implicit data, which can improve the robustness on the noise effectively. Mathematically, LatLRR is to minimize the following constrained objective function

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} rank(\mathbf{Z}_{O|H}) + rank(\mathbf{L}_{O|H}), \\ &s.t. \ \mathbf{X}_O = \mathbf{X}_O\mathbf{Z} + \mathbf{LX}_O, \end{aligned} \tag{5}$$

where $\mathbf{X}_O$ denotes the explicit data, and $\mathbf{L} \in \mathbb{R}^{m\times m}$ is used to extract "salient feature". For an arbitrary testing data vector $\mathbf{x} \in \mathbb{R}^{m\times 1}$, the transformed feature vector $\mathbf{y} \in \mathbb{R}^{m\times 1}$ can be obtained by $\mathbf{y} = \mathbf{Lx}$. Since rank function is non-convex and its optimization is an NP-hard issue, nuclear norm is a widely-used alternative to approximate the rank function. Besides, by taking noise into consideration, the LatLRR model can be reformulated as

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1, \\ &s.t. \ \mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E}. \end{aligned} \tag{6}$$

The LatLRR model is an extension of LRR, which can extract the implicit structure of data for subspace representation. LatLRR is also widely applied in realistic scenarios since it can not only reconstruct data correctly but also extract "salient features". Based on LatLRR, a series of methods have been extended, including the Frobenius norm-based method [45], iterative reweighted Frobenius norm-based method [46], and weighted Schatten p norm-based method [47]. However, there are two main disadvantages for the application of LatLRR: 1) The features extracted by LatLRR have the same dimension with the original data. 2) The model optimization involves the minimization of two nuclear norm functions, which is time consuming. So, by considering the computational complexity and memory consumption, it may be unreasonable for dealing with high-dimensional data directly using LatLRR.

Table I illustrates the models involved in the related works. To effectively explore the relationships between samples, LRR [38] has been proposed to mine the structure of data. Furthermore, some studies [42], [43], [48] have combined LRR and projection learning for better handling high-dimensional data. Although reducing computational complexity, these methods still ignore the significance of exploring feature correlations. To address this problem, the latent low-rank representation [44], [45] has been introduced to explore salient features in the data, which can effectively uncover the feature correlation of the data, but the scalability to handle high-dimensional data remains limited. To simultaneously explore the sample relationships and feature correlations among high-dimensional data, we propose the LatLRE for hyperspectral classification in Section III, which can fully consider the information from

TABLE I
COMPARISON OF VARIOUS MODELS IN RELATED WORKS.

Method	Objective function	Constraint	Low-rank representation	Projection learning	Latent low-rank representation
LRR [38]	$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1$	$\mathbf{X} = \mathbf{XZ} + \mathbf{E}$	✓		
LRE [42]	$\min_{\mathbf{Z},\mathbf{P}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1}$	$\mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{XZ} - \mathbf{E},\ \mathbf{P}^T\mathbf{P} = \mathbf{I}$	✓	✓	
LRSPP [43]	$\min_{\mathbf{P},\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \alpha \|\mathbf{Z}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda tr(\mathbf{P}^T\mathbf{XMX}^T\mathbf{P})$	$\mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{XZ} + \mathbf{E}$	✓	✓	
LRR_NP [48]	$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \frac{\beta}{2} tr(\mathbf{Z}(\mathbf{I}-\mathbf{W})^T(\mathbf{I}-\mathbf{W})\mathbf{Z}^T)$	$\mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{XZ} + \mathbf{E}$	✓	✓	
LatLRR [44]	$\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1$	$\mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E}$	✓		✓
FLLRR [45]	$\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_F + \|\mathbf{L}\|_F + \lambda \|\mathbf{E}\|_1$	$\mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E}$	✓		✓
Proposed LatLRE	$\min_{\mathbf{Z},\mathbf{L},\mathbf{E},\mathbf{P}} \|\mathbf{Z}\|_* + \alpha \|\mathbf{L}\|_L^1 + \lambda \|\mathbf{E}\|_{2,1}$	$\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}, \mathbf{P}^T\mathbf{P} = \mathbf{I}$	✓	✓	✓

multiple directions to achieve robust and efficient feature dimensionality reduction and image classification.

III. PROPOSED METHOD

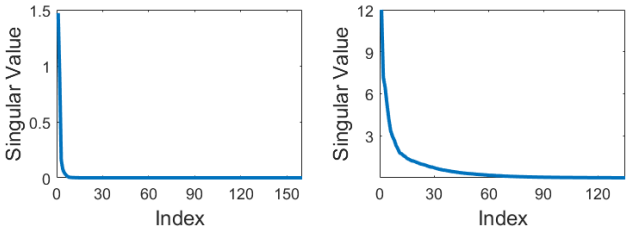
A. The Motivation and Objective Function

To mitigate the curse of dimensionality and more comprehensively exploit the implicit information, we propose latent low-rank embedding (LatLRE) by combining latent subspace representation with low-dimensional projection learning, which can maintain the robustness of the latent low-rank representation and simultaneously achieve dimensionality reduction. Given a HSIs dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{m\times N}$, to reduce the dimensionality of samples from m to d ($d \ll m$) without affecting the recognition performance, we formulate the following model

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} rank(\mathbf{Z}) + rank(\mathbf{L}) + \lambda \|\mathbf{E}\|_1, \\ &s.t.\ \mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}, \mathbf{P}^T\mathbf{P} = \mathbf{I}, \end{aligned} \tag{7}$$

where $\mathbf{Z}$, $\mathbf{L}$, $\mathbf{E}$, and λ represent reconstruction matrix, salient feature extraction matrix, noise matrix, and regularization parameter, respectively. In the latent subspace representation model (7), the original data matrix $\mathbf{X} \in \mathbb{R}^{m\times N}$ is reconstructed from column ($\mathbf{XZ}$) and row ($\mathbf{LX}$) according to the $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}$. Specifically, $\mathbf{Z} \in \mathbb{R}^{N\times N}$ reconstructs the data from the sample dimension, while $\mathbf{L} \in \mathbb{R}^{m\times m}$ is utilized to extract "salient features" and reconstruct the data $\mathbf{X}$ from the feature dimension. Taking Indian Pines dataset as a sample, we show the singular value curves of sample relationships $\mathbf{Z}$ and feature correlations $\mathbf{L}$ in Fig. 2, both of which drop rapidly and approach 0. Therefore, $rank(\cdot)$ is introduced to characterize their low-rankness. Compared with the existing low-rank feature dimensionality reduction methods, such as LRE [42] and latent low-rank and sparse embedding (LLRSE) [35], the LatLRE considers the information from both row and column directions rather than single direction. In this way, more global structure information can be captured for projection learning. Therefore, the robustness to complex noise and recognition accuracy can be improved.

As discussed before, due to ℓ_1-norm cannot characterize the geometric structure of data well, the noise matrix $\mathbf{E}$ measured by ℓ_1-norm usually fails to gain the expected performance



(a) (b)

Fig. 2. The low-rankness of (a) sample relationships $\mathbf{Z}$, and (b) feature correlations $\mathbf{L}$ on Indian Pines dataset.

[21], [42]. Moreover, although ℓ_1-norm can reduce the negative impact of outliers with high probability, it is still unclear whether ℓ_1-norm works on enhancing the role of small distance between data points from different classes. To this end, $\ell_{2,1}$-norm is considered to replace ℓ_1-norm for measuring the noise matrix $\mathbf{E}$. On the one hand, $\ell_{2,1}$-norm is robust to noise and does not destory the data structures. On the other hand, $\ell_{2,1}$-norm based methods are rotationally invariant. Thus, the model (7) can be reformulated as the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{L},\mathbf{E},\mathbf{P}} rank(\mathbf{Z}) + rank(\mathbf{L}) + \lambda \|\mathbf{E}\|_{2,1}, \\ &s.t.\ \mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E},\ \mathbf{P}^T\mathbf{P} = \mathbf{I}. \end{aligned} \tag{8}$$

In Eq. (8), each element $\mathbf{z}_{ij}$ in the coefficient matrix $\mathbf{Z}$ can be viewed as a representation of the relationships between each data pair $\mathbf{x}_i$ and $\mathbf{x}_j$. In other words, each column $\mathbf{z}_i$ in matrix $\mathbf{Z}$ represents the relationship between the ith sample and all N samples in the original data $\mathbf{X}$. Furthermore, the relationship between samples can reflect the structure of the data. Thus, we can infer that an ideal $\mathbf{Z}$ should possess the following two characteristics: 1) sparsity, ensuring the presence of as few redundant values as possible in $\mathbf{Z}$; 2) low rankness, allowing the relationships between samples to be characterized by fewer dominant components. For the reasons mentioned above, we impose a nuclear norm on $\mathbf{Z}$, resulting in the transformation of Eq. (8) to:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{L},\mathbf{E},\mathbf{P}} \|\mathbf{Z}\|_* + rank(\mathbf{L}) + \lambda \|\mathbf{E}\|_{2,1}, \\ &s.t.\ \ \mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E},\ \mathbf{P}^T\mathbf{P} = \mathbf{I}. \end{aligned} \tag{9}$$

Eq. (9) relaxes the rank constraint imposed on $\mathbf{Z}$ in Eq. (8) by using the nuclear norm. The nuclear norm is denoted by a sum of non-zero singular values. Typically, larger singular values carry more significant information compared to smaller singular values. This implies that larger singular values can

better characterize the primary relationships within the data. Note that nuclear norm minimization achieves singular value shrinkage through soft thresholding, where all singular values are equally subtracted from a common threshold β, as shown in Eq. (21) and Eq. (22). In this way, larger singular values are retained while small singular values are discarded to ensure low rank. Consequently, the major parts with richer characterization of the global relationships are preserved, allowing the global structure to be revealed more clearly.

Furthermore, differently from $\mathbf{Z}$ which depicts the sample relationships, $\mathbf{L}$ is used to depict the relationships between features. Specifically, each column $\mathbf{l}_i$ of $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_m]$ represents the relationships between the ith feature and all m features of the original data, which can reflect the correlation between features. And the special nature of HSIs, i.e., the continuity of the bands, thus reflecting the high similarity of neighbouring features, leads to more redundancy. Therefore, the challenge of efficiently handling highly redundant features while retaining reliable correlation between them cannot be ignored. In other words, a feasible solution to the above challenge is to use a suitable constraint on $\mathbf{L}$ to make the salient features of the original image more prominent by maintaining the sparsity while dealing with the relationships between the features in a more effective manner. Some recent works [49], [50] utilized logarithmic function to enforce a more precise rank surrogate, as the logarithmic norm has been demonstrated to provide a superior sparsity-driven surrogate for the rank function. Based on this, we introduce Definition 1 and use the matrix logarithmic norm as a rank surrogate.

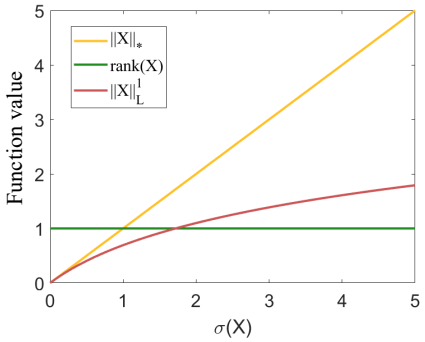


Fig. 3. Visualization of the rank function, the nuclear function, as well as the logarithmic function.

Definition 1 (Matrix Logarithmic Norm): [49] For any matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$, its matrix logarithmic norm with $0 < p \leq 1$ and $\epsilon > 0$ is defined as

$$\|\mathbf{X}\|_L^p = \sum_{i=1}^{min\{m,N\}} log(\sigma_i^p(\mathbf{X}) + \epsilon), \tag{10}$$

where $\sigma_i^p(\mathbf{X})$ denotes the ith singular value of $\mathbf{X}$.

From Eq. (10), it can be observed that unlike the nuclear norm, which simply sums all singular values, the matrix logarithmic norm sums the logarithms of all singular values. For a more intuitive presentation, Fig. 3 presents a graphical representation that offers a visual comparison of the rank function, the nuclear function, as well as the logarithmic function. From the Fig. 3, two observations can be made: 1) The nuclear norm more closely approximates the rank function when the singular value is small, while the logarithmic norm more closely approximates the rank function in the large singular value case. 2) The logarithmization of singular values will reduce the impact of large singular values, making the impact of larger singular values more equal, while exacerbating the impact of small singular values. Taking into account the points mentioned above, we introduce a logarithmic norm for the matrix $\mathbf{L}$ that more accurately approximates the rank constraint and thereby enabling a more accurate measurement of correlation. Moreover, building upon Definition 1, and we set $p = 1$ for convenience, the LatLRE model is

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{L},\mathbf{E},\mathbf{P}} \ & \|\mathbf{Z}\|_* + \alpha\|\mathbf{L}\|_L^1 + \lambda\|\mathbf{E}\|_{2,1}, \\ s.t. \ & \mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{X}\mathbf{Z} + \mathbf{L}\mathbf{X}) + \mathbf{E}, \mathbf{P}^T\mathbf{P} = \mathbf{I}, \end{aligned} \tag{11}$$

where α and λ represent the balance parameters. The LatLRE achieves sparsity in $\mathbf{L}$ by applying the logarithmic norm, effectively penalizing singular values approaching zero. Moreover, applying smaller and equalising penalties to larger singular values preserves important feature relationships and provides a more accurate measure of correlation while removing redundancy values.

Overall, the LatLRE allows a more comprehensive depiction of feature relationships from multiple directions and enables robust feature dimensionality reduction, as shown in Fig. 1. Specifically, LatLRE consists of the following components:

- The first term $\|\mathbf{Z}\|_*$ is introduced to characterize the low-rankness of sample relationships $\mathbf{Z}$. In this way, the relationships of the samples are preserved, so that the global structure is more clearly revealed.
- The second term $\|\mathbf{L}\|_L^1$ is used to describe the low-rankness of feature correlations $\mathbf{L}$. This way reduces redundant feature information and allows feature correlations to be exploited more effectively.
- The third term $\|\mathbf{E}\|_{2,1}$ is applied to constrain the sparsity of the noise $\mathbf{E}$ for ensuring the robustness of LatLRE.
- The constraint term $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{X}\mathbf{Z}+\mathbf{L}\mathbf{X})+\mathbf{E}$ decomposes the original data into principal feature representation $\mathbf{X}\mathbf{Z}$, salient feature representation $\mathbf{L}\mathbf{X}$ as well as noise matrix $\mathbf{E}$. And a projection matrix $\mathbf{P}$ is introduced to achieve feature dimensionality reduction.
- The constraint term $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ denotes an orthogonal constraint imposed on the projection matrix $\mathbf{P}$, which enables the projection matrix to be less correlated and reduces redundant information.

B. The Optimization Solution

In this section, a detailed description of the optimization process for the proposed LatLRE is provided. Specifically, we solve model (11) by alternating direction method of multipliers algorithm (ADMM) [51], [52], [53]. And an auxiliary variable $\mathbf{J}$ is first introduced by setting $\mathbf{Z} = \mathbf{J}$. Thus, the LatLRE model can be rewritten as

$$\begin{aligned} \min_{\mathbf{J},\mathbf{Z},\mathbf{L},\mathbf{E},\mathbf{P}} \ & \|\mathbf{J}\|_* + \alpha\|\mathbf{L}\|_L^1 + \lambda\|\mathbf{E}\|_{2,1}, \\ s.t. \ & \mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{X}\mathbf{Z} + \mathbf{L}\mathbf{X}) + \mathbf{E}, \mathbf{Z} = \mathbf{J}, \mathbf{P}^T\mathbf{P} = \mathbf{I}. \end{aligned} \tag{12}$$

Then, Eq. (12) can be solved by the ADMM, in which the underlying variables are updated alternately. The corresponding augmented Lagrangian function is

$$\begin{cases} \mathbf{J}_{k+1} & = \underset{\mathbf{J}}{argmin}\, \alpha\|\mathbf{J}\|_* + \frac{1}{2}\|\mathbf{J} - (\mathbf{Z} + \frac{\mathbf{W_2}}{\mu})\|_F^2, \\ \mathbf{Z}_{k+1} & = \underset{\mathbf{Z}}{argmin}\, \frac{\mu}{2}(\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX} - \mathbf{E}) + \frac{\mathbf{W_1}}{\mu}\|_F^2 + \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{W_2}}{\mu}\|_F^2), \\ \mathbf{L}_{k+1} & = \underset{\mathbf{L}}{argmin}\, \alpha\|\mathbf{L}\|_L^1 + \frac{\mu}{2}\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E} + \frac{\mathbf{W_1}}{\mu}\|_F^2, \\ \mathbf{E}_{k+1} & = \underset{\mathbf{E}}{argmin}\, \frac{\lambda}{\mu}\|\mathbf{E}\|_{2,1} + \frac{1}{2}\|\mathbf{E} - (\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \frac{\mathbf{W_1}}{\mu})\|_F^2, \\ \mathbf{P}_{k+1} & = \underset{\mathbf{P}}{argmin}\, \frac{\mu}{2}\|\mathbf{P}^T(\mathbf{X} - (\mathbf{XZ} + \mathbf{LX})) - \mathbf{E} + \frac{\mathbf{Y_1}}{\mu}\|_F^2, \\ \mathbf{W}_{k+1}^1 & = \mathbf{W}_k^1 + \mu(\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E}), \\ \mathbf{W}_{k+1}^2 & = \mathbf{W}_k^2 + \mu(\mathbf{Z} - \mathbf{J}). \end{cases} \tag{15}$$

$$\begin{aligned} &\mathcal{L}(\mathbf{P}, \mathbf{Z}, \mathbf{J}, \mathbf{L}, \mathbf{E}, \mathbf{W}^1, \mathbf{W}^2) \\ &= \|\mathbf{J}\|_* + \alpha\|\mathbf{L}\|_L^1 + \lambda\|\mathbf{E}\|_{2,1} \\ &+ \langle \mathbf{W}_1, \mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E}\rangle + \langle \mathbf{W}_2, \mathbf{Z} - \mathbf{J}\rangle \\ &+ \frac{\mu}{2}(\left\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E}\right\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2), \end{aligned} \tag{13}$$

where $\mathbf{W}^1, \mathbf{W}^2$ are the augmented Lagrangian multipliers, $\alpha > 0$ is a balance parameter, and $\mu > 0$ is the penalty parameter. By simple algebraic operation, the problem (13) can be rewritten as

$$\begin{aligned} &\mathcal{L}(\mathbf{P}, \mathbf{Z}, \mathbf{J}, \mathbf{L}, \mathbf{E}, \mathbf{W^1}, \mathbf{W^2}) \\ =&\|\mathbf{J}\|_* + \alpha\|\mathbf{L}\|_L^1 + \lambda\|\mathbf{E}\|_{2,1} \\ &+ \frac{\mu}{2}(\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E} + \frac{\mathbf{W^1}}{\mu}\|_F^2 \\ &+ \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{W^2}}{\mu}\|_F^2). \end{aligned} \tag{14}$$

Let $\mathbf{P}_{k+1}$, $\mathbf{Z}_{k+1}$, $\mathbf{J}_{k+1}$, $\mathbf{L}_{k+1}$, $\mathbf{E}_{k+1}$, $\mathbf{W}_{k+1}^1$, $\mathbf{W}_{k+1}^2$ denote the optimal variables and Lagrangian multipliers at the $(k+1)$-th iteration $(k = 1, 2, \cdots)$. Then, the overall optimization process is shown in Eq.(15). Moreover, the detailed updating schemes are formulated as follows.

Step.1: $[\mathbf{Update\ P}]$ By fixing all other variables to solve $\mathbf{P}$, the objective function is converted to a classic orthogonal procrustes problem expressed as

$$\begin{aligned} \mathbf{P}_{k+1} &= \underset{\mathbf{P}}{argmin}\, \frac{\mu}{2}\|\mathbf{P}^T(\mathbf{X} - (\mathbf{XZ} + \mathbf{LX})) - \mathbf{E} + \frac{\mathbf{W_1}}{\mu}\|_F^2 \\ &= \underset{\mathbf{P}}{argmin}\, \frac{\mu}{2}\|\mathbf{P}^T\mathbf{A} - \mathbf{B}\|_F^2, \end{aligned} \tag{16}$$

where $\mathbf{A} = \mathbf{X} - \mathbf{XZ} - \mathbf{LZ}$, $\mathbf{B} = \mathbf{E} - \mathbf{Y_1}/\mu$. The detail of orthogonal Procrustes problem is a matrix approximation problem in linear algebra and can be solved by the SVD. By denoting the SVD of $\mathbf{BA}^T$ as

$$\mathbf{BA}^T = (\mathbf{E} - \mathbf{Y_1}/\mu)(\mathbf{X} - \mathbf{XZ} - \mathbf{LX})^T = \mathbf{U\Sigma V}^T, \tag{17}$$

the optimal $\mathbf{P}$ can be updated by $\mathbf{P}_{k+1} = \mathbf{VU}^T$.

Step.2: $[\mathbf{Update\ Z}]$ Fix the variables other than $\mathbf{Z}$, problem (13) can be rewritten as

$$\begin{aligned} \mathbf{Z}_{k+1} =& \underset{\mathbf{Z}}{argmin}\, \frac{\mu}{2}(\|\mathbf{P}^T(\mathbf{X} - \mathbf{XZ} - \mathbf{LX}) - \mathbf{E} + \frac{\mathbf{W^1}}{\mu}\|_F^2 \\ &+ \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{W^2}}{\mu}\|_F^2). \end{aligned} \tag{18}$$

By calculating the derivative of $\mathbf{Z}$ in (6) and setting it to zero, we can obtain

$$\begin{aligned} \frac{\partial\mathcal{L}(\mathbf{Z})}{\partial\mathbf{Z}} =& \frac{\mu}{2}[2(\mathbf{Z} - \mathbf{J} + \frac{\mathbf{W}^2}{\mu}) + 2(\mathbf{X}^T\mathbf{PP}^T(\mathbf{X} \\ &-\mathbf{XZ} - \mathbf{LX}) - \mathbf{X}^T\mathbf{P}(\mathbf{E} - \frac{1}{\mu}\mathbf{W^1}))] = 0. \end{aligned} \tag{19}$$

Then, solving the above linear equation, it can obtain the optimal $\mathbf{Z}_{k+1}$ as

$$\begin{aligned} \mathbf{Z}_{k+1} = &(\mathbf{X}^T\mathbf{PP}^T\mathbf{X} + \mathbf{I}_1)^{-1} \times [\mathbf{X}^T\mathbf{P}(\mathbf{P}^T\mathbf{X} - \mathbf{P}^T\mathbf{LX} \\ &+ \frac{1}{\mu}\mathbf{W_1} - \mathbf{E}) + \mathbf{J} - \frac{1}{\mu}\mathbf{W^2}], \end{aligned} \tag{20}$$

where the $(\mathbf{X}^T\mathbf{PP}^T\mathbf{X} + \mathbf{I_1})^{-1}$ can be computed before iterations, and $\mathbf{I_1}$ is an identity matrix of size $N \times N$.

Step.3: $[\mathbf{Update\ J}]$ In this step, except $\mathbf{J}$, other variables are fixed. Thus, the problem (13) becomes

$$\mathbf{J}_{k+1} = \underset{\mathbf{J}}{argmin}\, \|\mathbf{J}\|_* + \frac{\mu}{2}\|\mathbf{J} - (\mathbf{Z} + \frac{\mathbf{W^2}}{\mu})\|_F^2, \tag{21}$$

which can be solved by the singular value thresholding (SVT) operator. And the optimal solution $\mathbf{J}_{k+1}$ is

$$\mathbf{J}_{k+1} = \mathcal{D}_{\frac{1}{\mu}}(\mathbf{Z} + \frac{\mathbf{W^2}}{\mu}), \tag{22}$$

where $\mathcal{D}_\beta(\mathbf{Y}) = \mathbf{U}_{M\times R}diag\{max(0, \sigma_i - \beta)\}\mathbf{V}_{N\times R}^T$ obeys

$$\mathcal{D}_\beta(\mathbf{Y}) = \underset{\mathbf{X}}{argmin}(\|\mathbf{X}\|_* + \frac{\mu}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2), \tag{23}$$

and σ_i is the i-th positive singular value of $\mathbf{Y}$, $1 \le i \le R$.

Step.4: $[\mathbf{Update\ L}]$ By regarding other variables as known to optimal $\mathbf{L}$, $\mathbf{L}_{k+1}$ denote the optimal variable and Lagrangian multiplier at the $(k+1)$-th iteration $(k = 0, 1, 2, ...)$. The problem becomes:

$$\begin{aligned} \mathbf{L}_{k+1} =& \underset{\mathbf{L}}{argmin}\, \alpha\|\mathbf{L}\|_L^1 \\ &+ \frac{\mu}{2}\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) - \mathbf{E} + \frac{\mathbf{W^1}}{\mu}\|_F^2. \end{aligned} \tag{24}$$

We set $\mathbf{M} = \mathbf{P}^T\mathbf{X} - \mathbf{P}^T\mathbf{XZ} - \mathbf{E} + \frac{\mathbf{W}^1}{\mu}$ for simplicity, and the updating of $\mathbf{L}_{k+1}$ can be written as

$$\mathbf{L}_{k+1} = \underset{\mathbf{L}}{argmin}\, \alpha\|\mathbf{L}\|_L^1 + \frac{\beta^k}{2}\|\mathbf{L} - \mathbf{L}_k + \frac{1}{\beta^k}d(\mathbf{L}_k)\|_F^2, \tag{25}$$

where $d(\mathbf{L}_k) = \mu\mathbf{P}(\mathbf{M} - \mathbf{P}^T\mathbf{L}_k\mathbf{X})\mathbf{X}^T$ represents the gradient of quadratic term. We use an acceleration scheme to enhance the convergence of algorithm, which can be expressed as:

$$\begin{cases} \mathbf{L}_k = \mathbf{L}_k + \omega_k(\mathbf{L}_k - \mathbf{L}_{k-1}), \\ \mathbf{L}_{k+1} = \underset{\mathbf{L}}{argmin}\, \alpha\|\mathbf{L}\|_L^1 + \frac{\beta^k}{2}\|\mathbf{L} - \mathbf{A}\|_F^2, \end{cases} \tag{26}$$

where $\mathbf{L}_k$ stands for the extrapolated point, ω_k represents an extrapolation weight constant, $\omega_{k+1} = (t_k - 1)/t_{k+1}$, $t_{k+1} = \frac{1}{2}(1+\sqrt{1+4{t_k}^2})$ and $\mathbf{A} = \mathbf{L}_k - \frac{1}{\beta^k}d(\mathbf{L}_k)$ is used for simplicity. The above problem can be resolved by obtaining an approximate expression in accordance with the following theorem:

Theorem 1 (Logarithmic Singular Value Thresholding [49]): For any $\mathbf{A} \in \mathbb{R}^{M\times N}$ and $\alpha > 0$, the SVD of $\mathbf{A}$ is $\mathbf{A} = \mathbf{U_A}\mathbf{\Sigma_A}\mathbf{V_A}$, the closed solution of the problem is

$$\mathbf{L}_{k+1} = \underset{\mathbf{X}}{argmin}\, \alpha\|\mathbf{X}\|_L^1 + \frac{1}{2}\|\mathbf{X}-\mathbf{A}\|_F^2, \tag{27}$$

where $\mathbf{X} = \mathbf{U_A}\Gamma_{\alpha,\epsilon}\mathbf{\Sigma_A}\mathbf{V_A}$ and the soft thresholding operator $\Gamma_{a,\epsilon}(\cdot)$ is defined as:

$$\begin{cases} \underset{a\in\{0,\frac{1}{2}(x-\epsilon+\sqrt{\Delta})\}}{argmin} h(a), & \Delta > 0, \\ 0, & \Delta \le 0, \end{cases} \tag{28}$$

where $\Delta = (x-\epsilon)^2 - 4(\alpha - x\epsilon)$ and function $h(a) = \frac{1}{2}(a-x)^2 + \alpha log(a+\epsilon), \mathbb{R}^+ \to \mathbb{R}^+$.

According to Theorem 1, we can obtain the optimal solution $\mathbf{L}_{k+1}$:

$$\mathbf{L}_{k+1} = \mathbf{U_A}\Gamma_{\frac{\alpha}{\beta^k},\epsilon}\mathbf{\Sigma_A}\mathbf{V_A}. \tag{29}$$

Step.5: [**Update E**] Let $\mathbf{M} = \mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ}+\mathbf{LX}) + \mathbf{W_1}/\mu$, $\mathbf{E}$ can be computed and updated by

$$\begin{aligned} \mathbf{E}_{k+1} =& \underset{\mathbf{E}}{argmin}\frac{\lambda}{\mu}\|\mathbf{E}\|_{2,1} \\ &+\frac{1}{2}\|\mathbf{E} - (\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ}+\mathbf{LX}) + \frac{\mathbf{W^1}}{\mu})\|_F^2 \\ =& \underset{\mathbf{E}}{argmin}\frac{\lambda}{\mu}\|\mathbf{E}\|_{2,1} + \frac{1}{2}\|\mathbf{E}-\mathbf{M}\|_F^2. \end{aligned} \tag{30}$$

From the view of vectors, Eq. (30) can be decomposed as

$$\underset{\mathbf{E}}{argmin}\frac{\lambda}{\mu}\|e_i\|_2 + \frac{1}{2}\|e_i - m_i\|_F^2, \tag{31}$$

where e_i and m_i are the i-th column of matrix $\mathbf{E}$ and $\mathbf{M}$, respectively. The solution to problem Eq. (30) is given by

$$e_i = \begin{cases} (1-\frac{\lambda/\mu}{\|m_i\|_2})m_i, & \|m_i\|_2 > \lambda/\mu, \\ 0, & \|m_i\|_2 \le \lambda/\mu. \end{cases} \tag{32}$$

Step.6: [**Update parameters**] The augmented Lagrangian multipliers $\mathbf{W}_1$ and $\mathbf{W}_2$ are updated by

$$\begin{aligned} \mathbf{W}_{k+1}^1 &\leftarrow \mathbf{W}^1 + \mu(\mathbf{P}^T(\mathbf{X}-\mathbf{XZ}-\mathbf{LX}) - \mathbf{E}), \\ \mathbf{W}_{k+1}^2 &\leftarrow \mathbf{W}^2 + \mu(\mathbf{Z}-\mathbf{J}), \end{aligned} \tag{33}$$

respectively, and penalty parameter μ is adjusted by

$$\mu = min(\rho\mu, \mu_{max}), \tag{34}$$

where $\rho > 1$ is a constant and μ_{max} is the upper bound of μ.

Finally, the complete ADMM algorithm for the proposed LatLRE model is summarized in **Algorithm 1**.

Algorithm 1 The LatLRE algorithm.

Input: Data $\mathbf{X}$, parameters $\lambda, \alpha, \mu_{min}$.

1: **Initialize:** $\mathbf{J}, \mathbf{Z}, \mathbf{L}, \mathbf{E}, \mathbf{W}^1, \mathbf{W}^2, k, t_1, \omega_1$, μ, μ_{max}, ρ, ε and $\mu^k = \mu_{min}$.
2: **while** no converge **do**
3: Update t_{k+1}, ω_{k+1}.
4: Update $\mathbf{P}, \mathbf{Z}, \mathbf{J}, \mathbf{L}, \mathbf{E}$.
5: Update the augmented Lagrangian multipliers $\mathbf{W}^1, \mathbf{W}^2$.
6: Update the penalty parameter μ.
7: Check the convergence conditions $\|\mathbf{P}^T\mathbf{X} - \mathbf{P}^T(\mathbf{XZ}+\mathbf{LZ}) - \mathbf{E}\|_\infty < \varepsilon$, and $\|\mathbf{Z}-\mathbf{J}\|_\infty < \varepsilon$.
8: **end while**

Output: Projection matrix $\mathbf{P}^T$.

C. Complexity Analysis

Given a HSIs dataset $\mathbf{X} \in \mathbb{R}^{m\times N}$, the major computational costs for LatLRE are the computation of $\mathbf{P} \in \mathbb{R}^{m\times d}$, $\mathbf{Z} \in \mathbb{R}^{N\times N}$, $\mathbf{J} \in \mathbb{R}^{N\times N}$, $\mathbf{L} \in \mathbb{R}^{m\times m}$, and $\mathbf{E} \in \mathbb{R}^{d\times N}$, where m is the dimensionality of each data point, N is the number of data points, and d is the reduced dimension. Specifically, for solving projection matrix $\mathbf{P}$, the main complexity comes from the SVD of matrix $\mathbf{BA}^T$. Since the size of $\mathbf{BA}^T$ is $m \times d$, the complexity of updating $\mathbf{P}$ is $O(d^3)$. Similarly, the major computational cost of $\mathbf{L}$ involves the computation of the SVD, which has a complexity of $O(m^3)$. The computational costs for solving $\mathbf{Z}$ is mainly spent on calculating the inverse of $(\mathbf{X}^T\mathbf{X} + \mathbf{I}_1)$, and the associated computational complexity is $O(N^3)$. For solving $\mathbf{J}$, the computational complexity is mainly contributed by singular value thresholding operator, the size of the input matrix $(\mathbf{Z}+\mathbf{W}^2/\mu)$ is $N \times N$. Thus, solving $\mathbf{J}$ costs about $O(N^3)$. And the computational complexity is $O(dN^2)$ for solving $\mathbf{E}$. In summary, assuming the LatLRE algorithm converges in k iterations. In practice, d ($d \ll m$ and $d \ll N$) is negligible after enlarging to the third power. Hence, the total computational complexity of LatLRE is about $O(kN^3)$.

D. Convergence Analysis

Due to the non-convexity of LatLRE and the difficulty of ensuring the strong convergence of the ADMM-based Algorithm 1, we aim to establish the weak convergence of our LatLRE algorithm by proving that the iterative sequence of Algorithm 1 converges to a stationary Karush-Kuhn-Tucker (KKT) point. Before proceeding with the proof, we first introduce the following theorem.

Theorem 2: For ADMM algorithm, let $\{\theta_k = (\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2)\}_{k=1}^\infty$ be the sequence generated by Algorithm 1, and the sequence $\{\theta\}_{k=1}^\infty$ is bounded.

Before proving the above theorem, we first introduce the following lemma.

Lemma 1: [54] Let $\mathcal{H}$ be a real Hilbert space endowed with an inner product $\langle\cdot,\cdot\rangle$, a norm $\|\cdot\|$, a dual norm $\|\cdot\|^\star$, and $y \in \partial\|x\|$, where $\partial f(x)$ is the subgradient of $f(x)$. Then $\|y\|^\star = 1$ if $x \neq 0$, and $\|y\|^\star \le 1$ if $x = 0$.

According to Lemma 1, the proof of Theorem 2 is as follows:

Proof of Theorem 2: Let $\mathbf{P}_{k+1}$, $\mathbf{Z}_{k+1}$, $\mathbf{J}_{k+1}$, $\mathbf{L}_{k+1}$, $\mathbf{E}_{k+1}$, $\mathbf{W}_{k+1}^1$, and $\mathbf{W}_{k+1}^2$ denote the optimal variables at the $(k+1)-th$ iteration. Next, we prove the boundedness of

the sequence $\{\theta\}_{k=1}^{\infty}$ in two parts: 1) prove that $\mathbf{W}_{k+1}^1$ and $\mathbf{W}_{k+1}^2$ are bounded; 2) prove that $\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}$, and $\mathbf{E}_{k+1}$ are bounded.

1) The boundedness of $\mathbf{W}_{k+1}^1$ and $\mathbf{W}_{k+1}^2$. According to the updating rule of $\mathbf{E}_k$, we have:

$$\begin{aligned}\mathbf{0} \in \lambda\partial\|\mathbf{E}_{k+1}\|_{2,1} + \mu(\mathbf{E}_{k+1} \\ - (\mathbf{P}_{k+1}^T\mathbf{X} - \mathbf{P}_{k+1}^T(\mathbf{X}\mathbf{Z}_{k+1} + \mathbf{L}_{k+1}\mathbf{X}) + \mathbf{W}_k^1/\mu),\end{aligned} \tag{35}$$

according to Eq.(33), i.e., $\mathbf{W}_{k+1}^1 = \mathbf{W}_k^1 + \mu(\mathbf{P}_{k+1}^T\mathbf{X} - \mathbf{P}_{k+1}^T(\mathbf{X}\mathbf{Z}_{k+1}+\mathbf{L}_{k+1}\mathbf{X})-\mathbf{E}_{k+1})$, Eq.(35) can be transformed into

$$\mathbf{0} \in \lambda\partial\|\mathbf{E}_{k+1}\|_{2,1} - \mathbf{W}_{k+1}^1. \tag{36}$$

Thus, we can obtain

$$\frac{1}{\lambda}w_{i\,k+1}^1 \in \partial\|e_{i\,k+1}\|_2, \tag{37}$$

where w_i and e_i are the ith column of $\mathbf{W}_{k+1}^1$ and $\mathbf{E}_{k+1}$, respectively. Introducing the Lemma 1 and based on the fact that the l_2 norm is self-dual [55], from Eq.(37), we can deduce that $\|w_{i\,k+1}^1\|_2 \leq \lambda$. Thus, the $\{\mathbf{W}_{k+1}^1\}$ is bounded. Next, we discuss the boundedness of the sequence $\{\mathbf{W}_{k+1}^2\}$. Since

$$\mathbf{0} \in \partial\|\mathbf{J}_{k+1}\|_* + \mu(\mathbf{J}_{k+1} - (\mathbf{Z}_{k+1}^T + \mathbf{W}_k^2/\mu), \tag{38}$$

and $\mathbf{W}_{k+1}^2 = \mathbf{W}_k^2 + \mu(\mathbf{Z}_{k+1} - \mathbf{J}_{k+1})$. Then, Eq.(38) can be transformed into

$$\mathbf{0} \in \partial\|\mathbf{J}_{k+1}\|_* - \mathbf{W}_{k+1}^2. \tag{39}$$

We can get

$$\mathbf{W}_{k+1}^2 \in \partial\|\mathbf{J}_{k+1}\|_*. \tag{40}$$

According to Lemma 1 and the fact that the dual norm of the nuclear norm $\|\cdot\|_*$ is the l_2 norm $\|\cdot\|_2$, and we have $\|\mathbf{W}_{k+1}^2\|_2 \leq 1$. Through the above derivation, it can be proven that $\{\mathbf{W}_{k+1}^2\}$ is bounded.

2) The boundedness of $\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}$, and $\mathbf{E}_{k+1}$. According to the updating rule of $\mathbf{L}_k$ in Algorithm 1, it can be naturally concluded that $F(\{\mathbf{L}_k\}) = \alpha\|\mathbf{L}_k\|_L^1 + \frac{\mu}{2}\|\mathbf{P}_k^T\mathbf{X} - \mathbf{P}_k^T(\mathbf{X}\mathbf{Z}_\mathbf{k} + \mathbf{L}_\mathbf{k}\mathbf{X}) - \mathbf{E}_\mathbf{k} + \mathbf{W}_k^1/\mu\|_F^2$ is non-increasing. And we can obtain that $F(\{\mathbf{L}_1\}) \geq F(\{\mathbf{L}_k\}) \geq \alpha\|\mathbf{L}_k\|_L^1$. Therefore, the boundedness of the sequence $\{\mathbf{L}_{k+1}\}$ can be held. Since $\mathbf{P}_{k+1}$ is imposed an orthogonal constraint, the boundedness of $\{\mathbf{P}_{k+1}\}$ can be held. Under the iterative rules of Algorithm 1, we can derive

$$\begin{aligned}&\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}, \mathbf{E}_{k+1}, \mathbf{W}_k^1, \mathbf{W}_k^2, \mu_k) \\ \leq&\mathcal{L}(\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2, \mu_k) \\ =&\mathcal{L}(\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_{k-1}^1, \mathbf{W}_{k-1}^2, \mu_{k-1}) \\ &+ \left\langle \mathbf{W}_k^1 - \mathbf{W}_{k-1}^1, \mathbf{P}_k^T\mathbf{X} - \mathbf{P}_k^T(\mathbf{X}\mathbf{Z}_k + \mathbf{L}\mathbf{X}) - \mathbf{E}_k \right\rangle \\ &+ \left\langle \mathbf{W}_k^2 - \mathbf{W}_{k-1}^2, \mathbf{Z}_k - \mathbf{J}_k \right\rangle \\ &+ \frac{\mu_k - \mu_{k-1}}{2}\left\|\mathbf{P}_k^T\mathbf{X} - \mathbf{P}_k^T(\mathbf{X}\mathbf{Z}_k + \mathbf{L}\mathbf{X}) - \mathbf{E}_k\right\|_F^2 \\ &+ \frac{\mu_k - \mu_{k-1}}{2}\left\|\mathbf{Z}_k - \mathbf{J}_k\right\|_F^2.\end{aligned} \tag{41}$$

And Eq.(41) can be transformed into

$$\begin{aligned}&\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}, \mathbf{E}_{k+1}, \mathbf{W}_k^1, \mathbf{W}_k^2, \mu_k) \\ \leq&\mathcal{L}(\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_{k-1}^1, \mathbf{W}_{k-1}^2, \mu_{k-1}) \\ &+ \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}(\left\|\mathbf{W}_k^1 - \mathbf{W}_{k-1}^1\right\|_F^2 + \left\|\mathbf{W}_k^2 - \mathbf{W}_{k-1}^2\right\|_F^2) \\ \leq&\mathcal{L}(\mathbf{P}_1, \mathbf{Z}_1, \mathbf{J}_1, \mathbf{L}_1, \mathbf{E}_1, \mathbf{W}_0^1, \mathbf{W}_0^2, \mu_0) \\ &+ \sum_k^n \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}(\left\|\mathbf{W}_k^1 - \mathbf{W}_{k-1}^1\right\|_F^2 + \left\|\mathbf{W}_k^2 - \mathbf{W}_{k-1}^2\right\|_F^2),\end{aligned} \tag{42}$$

where n denotes the number of iterations. According to Eq.(34), i.e., $\mu_k = \rho^k\mu_0$, it is easy to get

$$\begin{aligned}\sum_k^n \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2} =& \sum_k^n \frac{\rho\mu_{k-1} + \mu_{k-1}}{2\mu_{k-1}^2} \\ \leq& \frac{\rho(\rho+1)}{2\mu_0(\rho-1)} \leq +\infty,\end{aligned} \tag{43}$$

so, $\sum_k^n \frac{\mu_k+\mu_{k-1}}{2\mu_{k-1}^2}$ is bounded. Since $\mathcal{L}(\mathbf{P}_1, \mathbf{Z}_1, \mathbf{J}_1, \mathbf{L}_1, \mathbf{E}_1, \mathbf{W}_0^1, \mathbf{W}_0^2, \mu_0)$ is finite, and $\sum_k^n \frac{\mu_k+\mu_{k-1}}{2\mu_{k-1}^2}$, $\mathbf{W}_k^1$, $\mathbf{W}_k^2$ are bounded, all terms on the right side of Eq. (42) is bounded, hence $\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}, \mathbf{E}_{k+1}, \mathbf{W}_k^1, \mathbf{W}_k^2, \mu_k)$ is bounded. Additionally, based on Algorithm 1, we can also obtain

$$\begin{aligned}&\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}_{k+1}, \mathbf{E}_{k+1}, \mathbf{W}_k^1, \mathbf{W}_k^2, \mu_k) \\ &+ \frac{1}{2\mu_k}(\left\|\mathbf{W}_k^1\right\|_F^2 + \left\|\mathbf{W}_k^2\right\|_F^2) \\ =&\|\mathbf{J}_{k+1}\|_* + \alpha\|\mathbf{L}_{k+1}\|_L^1 + \lambda\|\mathbf{E}_{k+1}\|_{2,1} \\ &+ \frac{\mu}{2}(\|\mathbf{P}_{k+1}^T\mathbf{X} - \mathbf{P}_{k+1}^T(\mathbf{X}\mathbf{Z}_{k+1} + \mathbf{L}_{k+1}\mathbf{X} - \mathbf{E}_{k+1}) + \frac{\mathbf{W}_k^1}{\mu_k}\|_F^2 \\ &+ \|\mathbf{Z}_{k+1} - \mathbf{J}_{k+1} + \frac{\mathbf{W}_k^2}{\mu_k}\|_F^2).\end{aligned} \tag{44}$$

It can be observed that in Eq. (44), the terms on the left side are bounded, and the boundedness of the terms on the right side is guaranteed. Therefore, $\{\mathbf{J}_{k+1}\}$, $\{\mathbf{E}_{k+1}\}$, and the last term $\|\mathbf{Z}_{k+1} - \mathbf{J}_{k+1} + \frac{\mathbf{W}_k^2}{\mu_k}\|_F^2$ are bounded. According to the above proof, i.e., $\{\mathbf{J}_{k+1}\}$, $\{\mathbf{W}_k^2\}$, and $\|\mathbf{Z}_{k+1} - \mathbf{J}_{k+1} + \frac{\mathbf{W}_k^2}{\mu_k}\|_F^2$ are bounded, it is easy to verify the boundedness of $\{\mathbf{Z}_{k+1}\}$. ■

In summary, we can conclude that $\{\theta_k = (\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2)\}_{k=1}^{\infty}$ is bounded. Subsequently, the convergence of LatLRE algorithm will be proved by the following theorem.

Theorem 3: The sequence $\{\theta_k\}_{k=1}^{\infty}$ is bounded, and assuming $\lim_{k\to\infty}\{\theta_{k+1} - \theta_k\} = 0$, then the accumulation point in the sequence $\{\theta_k\}_{k=1}^{\infty}$ satisfies the KKT conditions. Specifically, when $\{\theta_k\}_{k=1}^{\infty}$ converges, it converges to a KKT point.

Proof of Theorem 3: Without loss of generality, we denote that the sequence $\{\theta_k\}_{k=1}^{\infty}$ is represented by itself, and it converges to an accumulation point $\theta_\star$, which can be given by $\lim_{k\to\infty}(\mathbf{P}_k, \mathbf{Z}_k, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2) =$

$(\mathbf{P}_\star, \mathbf{Z}_\star, \mathbf{J}_\star, \mathbf{L}_\star, \mathbf{E}_\star, \mathbf{W}_\star^1, \mathbf{W}_\star^2)$. Considering the boundedness and the update rules of $\mathbf{W}_k^1$ and $\mathbf{W}_k^2$, we have

$$\mathbf{P}_\star^T\mathbf{X} - \mathbf{P}_\star^T(\mathbf{X}\mathbf{Z}_\star + \mathbf{L}_\star\mathbf{X}) - \mathbf{E}_\star = \mathbf{0}, \tag{45}$$

and

$$\mathbf{Z}_\star - \mathbf{J}_\star = \mathbf{0}. \tag{46}$$

Taking the updating rules of $\mathbf{Z}_k$, we can obtain

$$\begin{aligned}
&\partial_{\mathbf{Z}}\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}, \mathbf{J}_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2)|_{\mathbf{Z}_{k+1}} \\
=&\mu_k(\mathbf{Z}_{k+1} - \mathbf{J}_k + \frac{\mathbf{W}_k^2}{\mu_k}) - \mu_k(\mathbf{X}^T\mathbf{P}_{k+1}(\mathbf{P}_{k+1}^T \\
\times&(\mathbf{X} - \mathbf{X}\mathbf{Z}_{k+1} - \mathbf{L}_k\mathbf{X}) - \mathbf{E}_k + \frac{\mathbf{W}_k^1}{\mu_k}) \\
=&\mu_k(\mathbf{Z}_{k+1} - \mathbf{J}_{k+1} + \frac{\mathbf{W}_k^2}{\mu_k}) + \mu_k(\mathbf{J}_{k+1} - \mathbf{J}_k) \\
-&\mu_k(\mathbf{X}^T\mathbf{P}_{k+1}(\mathbf{P}_{k+1}^T(\mathbf{X} - \mathbf{X}\mathbf{Z}_{k+1} - \mathbf{L}_{k+1}\mathbf{X}) - \mathbf{E}_{k+1} + \frac{\mathbf{W}_k^1}{\mu_k}) \\
+&\mu_k(\mathbf{X}^T\mathbf{P}_{k+1}(\mathbf{L}_{k+1} - \mathbf{L}_k)\mathbf{X} + \mathbf{X}^T\mathbf{P}_{k+1}(\mathbf{E}_{k+1} - \mathbf{E}_k)) \\
=&\mathbf{W}_k^2 - \mathbf{X}^T\mathbf{P}_{k+1}\mathbf{W}_k^2 + \mu_k\mathbf{X}^T\mathbf{P}_{k+1} \\
\times&((\mathbf{L}_{k+1} - \mathbf{L}_k)\mathbf{X} + (\mathbf{E}_{k+1} - \mathbf{E}_k)) = 0.
\end{aligned} \tag{47}$$

Assuming $\lim_{k\to\infty}(\mathbf{J}_{k+1} - \mathbf{J}_k) = 0$, $\lim_{k\to\infty}(\mathbf{L}_{k+1} - \mathbf{L}_k) = 0$, and $\lim_{k\to\infty}(\mathbf{E}_{k+1} - \mathbf{E}_k) = 0$, then $\mathbf{W}_k^2 - \mathbf{X}^T\mathbf{P}_{k+1}\mathbf{W}_k^2 = 0$ can hold. In this case, through Eq. (47), we can get

$$\begin{aligned}
\mathbf{0} &\in \partial_{\mathbf{Z}}\mathcal{L}(\mathbf{P}_\star, \mathbf{Z}_\star, \mathbf{J}_\star, \mathbf{L}_\star, \mathbf{E}_\star, \mathbf{W}_\star^1, \mathbf{W}_\star^2) \\
&= \mathbf{W}_\star^2 - \mathbf{X}^T\mathbf{P}_\star\mathbf{W}_\star^2.
\end{aligned} \tag{48}$$

From Eq. (39), we have

$$\begin{aligned}
\mathbf{0} &\in \partial_{\mathbf{J}}\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}, \mathbf{L}_k, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2)|_{\mathbf{J}_{k+1}} \\
&= \partial\|\mathbf{J}_{k+1}\|_* - \mathbf{W}_{k+1}^2.
\end{aligned} \tag{49}$$

When $k \to \infty$, we can get

$$\mathbf{0} \in \partial_{\mathbf{J}}\mathcal{L}(\mathbf{P}_\star, \mathbf{Z}_\star, \mathbf{J}_\star, \mathbf{L}_\star, \mathbf{E}_\star, \mathbf{W}_\star^1, \mathbf{W}_\star^2). \tag{50}$$

Furthermore, we can deduce the following equation:

$$\begin{aligned}
\mathbf{0} &\in \partial_{\mathbf{L}}\mathcal{L}(\mathbf{P}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{J}_{k+1}, \mathbf{L}, \mathbf{E}_k, \mathbf{W}_k^1, \mathbf{W}_k^2)|_{\mathbf{L}_{k+1}} \\
&= \alpha\partial\|\mathbf{L}_{k+1}\|_L - \mu_k\mathbf{P}_{k+1}(\mathbf{P}_{k+1}^T(\mathbf{X} - \mathbf{X}\mathbf{Z}_{k+1} \\
&\quad + \mathbf{L}_{k+1}\mathbf{X}) - \mathbf{E}_k + \frac{\mathbf{W}_k^1}{\mu_k})\mathbf{X}^T \\
&= \alpha\partial\|\mathbf{L}_{k+1}\|_L - \mathbf{P}_{k+1}\mathbf{W}_{k+1}^1\mathbf{X}^T \\
&\quad + \mu_k\mathbf{P}_{k+1}(\mathbf{E}_{k+1} - \mathbf{E}_k)\mathbf{X}^T.
\end{aligned} \tag{51}$$

Similarly, if $\lim_{k\to\infty}(\mathbf{E}_{k+1} - \mathbf{E}_k) = 0$, we can get

$$\mathbf{0} \in \partial_{\mathbf{L}}\mathcal{L}(\mathbf{P}_\star, \mathbf{Z}_\star, \mathbf{J}_\star, \mathbf{L}_\star, \mathbf{E}_\star, \mathbf{W}_\star^1, \mathbf{W}_\star^2). \tag{52}$$

From Eq. (36), the following equation holds:

$$\mathbf{0} \in \partial_{\mathbf{E}}\mathcal{L}(\mathbf{P}_\star, \mathbf{Z}_\star, \mathbf{J}_\star, \mathbf{L}_\star, \mathbf{E}_\star, \mathbf{W}_\star^1, \mathbf{W}_\star^2). \tag{53}$$

In summary, according to Eq. (45), Eq. (46), Eq. (48), Eq. (50), Eq. (52), and Eq. (53), we can prove that the accumulation point $\theta_\star$ of the sequence $\{\theta_k\}_{k=1}^\infty$ satisfies the KKT conditions. ■

In summary, the convergence of the proposed algorithm is proved by Theorem 3.

IV. Experimental Results and Analysis

In this section, experiments are conducted for verifying the effectiveness of LatLRE.

A. Hyperspectral Datasets

1) Indian Pines (IP) dataset: The IP dataset was obtained by the AVIRIS sensor in 1992. The IP dataset contains images with a spatial dimension of 145×145 and a resolution of 20 m/pixel, with 224 spectral bands covering a wavelength range of 400 to 2500 nm. In the experiments, 200 spectral bands are used. And only 50% (10,249) of the total 21,025 pixels include ground truth information from the 16 different land-cover classes. In addition, we randomly select 10% of the hyperspectral data as the training set and the rest for testing.

TABLE II
Classification Performance of Different FDR Methods on Indian Pine Dataset.

Class	Method						
	NPE	LPP	LRE	RPL	S^3PCA	GRSC	LatLRE
1	31.71	17.07	53.66	53.66	**63.41**	43.90	53.66
2	75.49	62.41	96.42	96.34	96.81	93.23	**98.37**
3	64.39	49.80	91.43	91.43	95.05	**96.92**	93.71
4	96.24	81.22	98.12	98.12	98.12	91.08	**99.06**
5	76.09	77.70	**96.78**	**96.78**	94.28	94.71	**93.56**
6	95.28	85.28	96.65	96.80	95.59	**97.56**	96.50
7	0	0	0	0	0	52.00	**84.00**
8	99.07	**99.30**	**99.30**	99.07	95.35	99.07	99.07
9	0	0	**72.22**	**72.22**	**72.22**	0	22.22
10	71.43	63.66	86.63	86.63	84.34	85.94	**93.83**
11	90.95	91.94	97.69	97.65	96.15	97.28	**98.14**
12	66.10	67.23	91.20	91.57	83.52	89.51	**94.38**
13	89.67	90.22	96.20	96.20	**98.91**	96.20	97.28
14	94.02	92.62	98.42	98.42	**99.03**	96.49	98.42
15	89.91	64.84	95.39	93.95	**96.54**	89.05	95.10
16	69.05	69.05	86.90	86.90	66.67	**89.29**	85.71
OA	82.89	78.02	94.91	94.86	93.88	93.95	**96.48**
AA	69.34	63.90	84.81	84.73	83.50	82.01	**87.89**
κ	80.35	74.73	94.19	94.13	93.01	93.09	**95.98**

2) Heihe dataset: The dataset was acquired through the CASI/SASI sensor located in the Zhangye Basin in the middle reaches of the Heihe River Basin, Gansu Province, China. The image contains 684 × 453 pixels with 135 bands after removing 14 bands heavily affected by noise. It comprises 8 classes, whereby 20 samples of each class are selected for training, and the remaining samples are reserved for testing.

3) WHU-Hi-HongHu dataset: This dataset was collected by Headwall Nano-Hyperspec sensors from Honghu, Hubei Province, China. And it comprises 270 bands with a wavelength range of 0.4-1 μm and a spatial resolution of 0.043m per pixel. For the experiment, a scene consists 940 × 475 pixels and contains 22 classes. Additionally, a training set consisting of 2% of the hyperspectral data is randomly selected, with the remainder reserved for testing.

B. Experimental Setup

To further validate the performance of LatLRE, several related unsupervised FDR methods, such as NPE [56], LPP [25], LRE [42], RPL [57], S^{3}PCA [24] and GRSC [58]

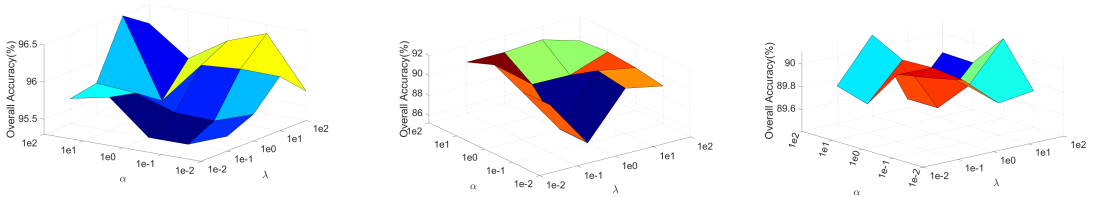


Fig. 4. The performance of LatLRE with respect to different parameters λ and α. (a) Indian Pines dataset. (b) Heihe dataset. (c) WHU-Hi-HongHu dataset.

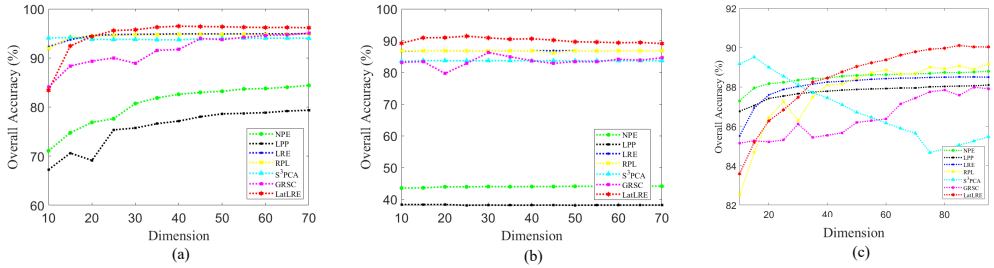


Fig. 5. The OA of different FDR methods with respect to different dimensions. (a) Indian Pines dataset. (b) Heihe dataset. (c) WHU-Hi-HongHu dataset.

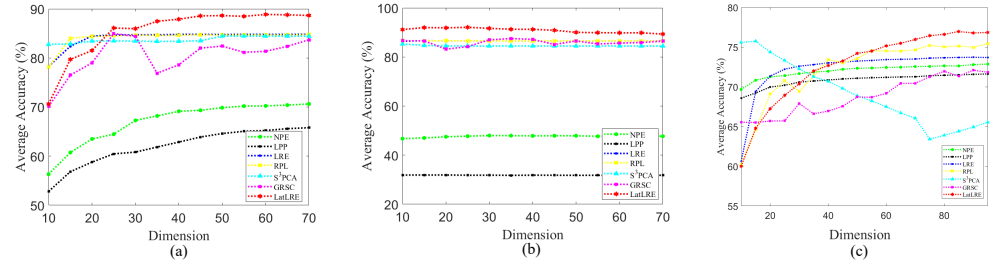


Fig. 6. The AA of different FDR methods with respect to different dimensions. (a) Indian Pines dataset. (b) Heihe dataset. (c) WHU-Hi-HongHu dataset.

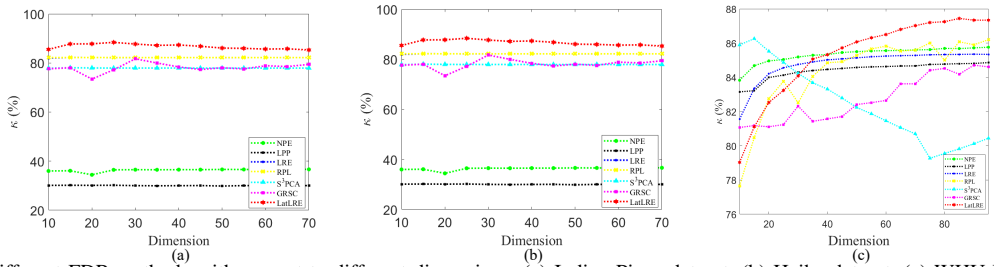


Fig. 7. The κ of different FDR methods with respect to different dimensions. (a) Indian Pines dataset. (b) Heihe dataset. (c) WHU-Hi-HongHu dataset.

are selected as comparative methods. Here, the class-specific accuracy, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) is utilized to evaluate the performances of different FDR methods and the SVM is adopted as classifier for recognition. The experimental platform is MATLAB R2016a.

C. Parameter Sensitivity Analysis

For LatLRE, there are two regularization parameters (i.e., α and λ) need to be determined. Fig. 4 shows the variation of recognition accuracies with different values of α and λ for three datasets. Taking the Indian Pines dataset as an example, through cross-validation experiments, α and λ are tuned in the range of {1e-2, 1e-1, 1e0, 1e1, 1e2}. It can be observed that {α=1e2, λ=1e0} can be set for the Indian Pines dataset. Similar to the above steps, the suitable parameters for Heihe dataset and WHU-Hi-HongHu dataset are {α=1e0, λ=1e-2} and {α=1e-1, λ=1e1}, respectively.

D. Implementation Details

Variable initialization: In the experiment, the proposed LatLRE method requires initialization for the variables $\mathbf{P}, \mathbf{Z}, \mathbf{J}, \mathbf{L}, \mathbf{E}, \mathbf{W}^1, \mathbf{W}^2$. Specifically, we perform a random initialization for $\mathbf{P}, \mathbf{Z}, \mathbf{J}$, and $\mathbf{L}$, and initialize $\mathbf{E}$, $\mathbf{W}^1$ and $\mathbf{W}^2$ with a zero matrix, respectively.

Parameter setting: Further, we need to define the parameters μ, μ_{max}, ρ, ε, t, ω, and ϵ with the settings $\mu = 10^{-5}$, $\mu_{max} = 10^7$, $\rho = 1.2$, $\varepsilon = 10^{-4}$, $t = 1.6$, $\epsilon = 0.1$, and $\omega = 1$.

Convergence conditions: During the experiments, we have set the maximum number of iterations to 200. Moreover, the convergence conditions for the LatLRE model have been set as follows:

$$\begin{cases} \|\mathbf{P}^T(\mathbf{X} - (\mathbf{XZ} + \mathbf{LX})) - \mathbf{E} + \frac{\mathbf{W^1}}{\mu}\|_F^2 & < \varepsilon, \\ \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{W^2}}{\mu}\|_F^2 & < \varepsilon. \end{cases} \tag{54}$$

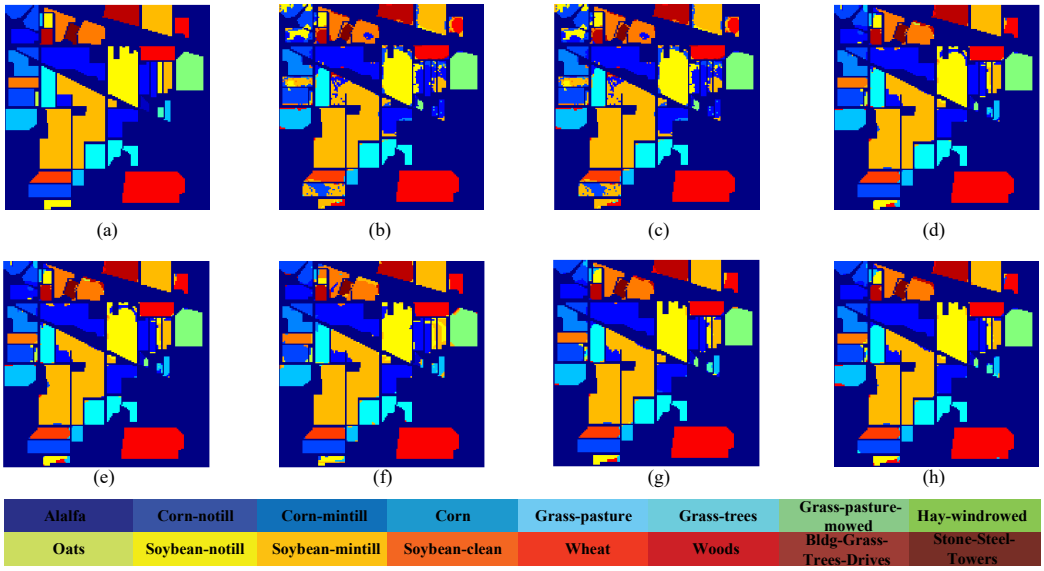


Fig. 8. Visualization of classification results by different methods on Indian Pines dataset. (a) Groundtruth of Indian Pines dataset. (b) NPE. (c) LPP. (d) LRE. (e) RPL. (f) S^3PCA. (g) GRSC. (h) LatLRE.

E. Analysis of Feature Dimensions

To explore the effectiveness of proposed feature dimensionality reduction (FDR) method LatLRE, we conduct numerical simulation experiments for sensitivity analysis of feature dimensions on the Indian Pines dataset, Heihe dataset as well as the WHU-Hi-HongHu dataset. To explore the impact of dimensionality on the performance of FDR models, we set the dimensions within the range of $\{10, 20, \cdots, 70\}$ on Indian Pines and Heihe datasets, which can be shown in Figs. 5-7 (a) and (b). Due to the large number of bands in the WHU-Hi-HongHu dataset, we chose the set of reduced dimensionality as $\{10, 20, \cdots, 95\}$, as shown in Figs. 5-7 (c). It can be found from the above experiments that our method has good results in reducing the feature dimensions on the three datasets. Furthermore, LatLRE achieves the best performance on the Indian Pines and Heihe dataset when the number of dimensionality reduction is 40 and 25, respectively. When conducting experiments on the WHU-Hi-HongHu dataset, we can find that the proposed method works best when the number of dimensionality reduction is 85.

F. Analysis and Discussions of Performance

We perform a comparative analysis of the proposed LatLRE with several traditional FDR methods, i.e., NPE [56], LPP [25], and LRE [42]. Additionally, we also evaluated the state-of-the-art methods such as RPL [57], S^{3}PCA [24] and GRSC [58]. The classification performances are presented in Tables II-IV. To facilitate comparison, the best result for each evaluation metric is highlighted in bold. And we further display the classification maps across the three datasets, as shown in Figs. 8-10. Through the experimental results, it is noted that the LatLRE is superior to the other comparative methods as it achieves the best results in terms of OA, AA, and κ on the three datasets. In particular, on the Indian Pines dataset, the LatLRE outperforms the other FDR methods

TABLE III
CLASSIFICATION PERFORMANCE OF DIFFERENT FDR METHODS ON HEIHE DATASET.

Class	Method						
	NPE	LPP	LRE	RPL	S^3PCA	GRSC	LatLRE
1	0	0	87.48	87.48	73.73	86.17	**92.83**
2	90.06	80.25	97.34	97.31	**99.25**	80.64	95.60
3	72.19	76.47	71.74	71.76	82.14	80.10	**83.19**
4	0	0	81.95	81.85	**89.11**	78.85	86.90
5	92.09	2.99	89.60	89.68	69.96	94.12	**95.43**
6	93.96	**94.68**	91.42	91.24	82.30	92.08	90.76
7	27.77	0.21	83.52	83.42	87.25	90.67	**92.95**
8	0	0.58	89.82	89.82	93.80	89.71	**99.06**
OA	43.60	38.34	86.83	86.82	83.72	83.40	**91.40**
AA	47.01	31.90	86.61	86.57	84.69	86.54	**92.09**
κ	36.09	30.17	82.23	82.21	78.06	78.05	**88.40**

on the OA, AA, and κ metrics by 1.57%-18.46%, 3.08%-23.19%, and 1.79%-21.25%, respectively. Several methods have demonstrated promising performance, such as LRE and RPL, which achieved satisfactory overall accuracy (OA) metrics across three datasets. However, these two methods exhibit a recognition accuracy of 0 for the class 7 in the Indian Pines dataset, which can also be clearly seen on Fig. 8. Similarly, we observe that S^3PCA and GRSC perform well on the Indian Pines dataset and WHU-Hi-HongHu dataset but fail to adapt effectively to the Heihe dataset. In contrast, the LatLRE not only demonstrates excellent performance on each individual class but also exhibits adaptability across all three aforementioned datasets. Based on the above observations we can conclude the following:

1) The superiority of LatLRE over S^3PCA is attributed to the introduction of projection learning. In contrast to PCA-based approaches, projection learning effectively preserves crucial structures within the data, mitigating the risk of overlooking vital information during the dimensionality reduction process inherent to PCA-based methods.

2) Methods such as LPP and RPL, which also employ

TABLE IV
CLASSIFICATION PERFORMANCE OF DIFFERENT FDR METHODS ON WHU-HI-HONGHU DATASET.

Class	Method						
	NPE	LPP	LRE	RPL	S^3PCA	GRSC	LatLRE
1	**96.45**	96.37	94.32	95.09	67.81	95.60	96.19
2	75.60	74.14	76.41	79.75	51.05	75.51	**80.59**
3	**94.76**	94.37	92.46	93.14	80.95	92.57	94.37
4	99.21	99.20	98.68	98.83	**99.84**	98.53	99.01
5	72.94	70.50	69.76	70.86	45.86	61.03	**73.61**
6	92.98	92.99	91.63	92.01	88.25	92.07	**93.15**
7	82.29	80.58	81.98	82.90	82.09	80.44	**84.18**
8	28.79	29.60	27.94	29.93	**29.98**	20.06	**33.38**
9	**96.53**	96.50	94.98	95.11	88.81	94.62	95.98
10	70.04	69.41	71.74	71.89	75.35	68.57	**75.50**
11	58.70	56.70	66.24	67.08	68.21	60.05	**70.79**
12	58.09	56.59	67.11	66.50	**76.36**	60.30	67.26
13	74.44	70.44	74.83	76.57	67.72	74.56	**78.13**
14	71.05	68.80	71.74	74.63	63.06	71.70	**76.31**
15	66.70	**70.98**	49.49	56.01	26.27	65.99	64.05
16	**93.78**	93.26	88.46	89.05	71.50	88.27	92.82
17	71.29	66.41	73.90	74.34	65.29	76.24	**76.27**
18	68.44	64.64	71.90	72.12	71.61	68.44	**72.28**
19	**87.22**	86.11	84.57	85.71	77.28	85.10	86.33
20	**80.77**	79.51	77.72	76.93	51.55	73.54	80.44
21	3.32	0.46	37.82	**45.66**	17.83	20.22	40.05
22	54.58	55.42	57.82	59.11	50.16	46.22	**62.31**
OA	88.73	88.05	88.51	89.07	85.05	87.58	**90.11**
AA	72.63	71.50	73.70	75.15	64.40	71.35	**76.96**
κ	85.68	84.80	85.34	86.08	79.82	84.18	**87.44**

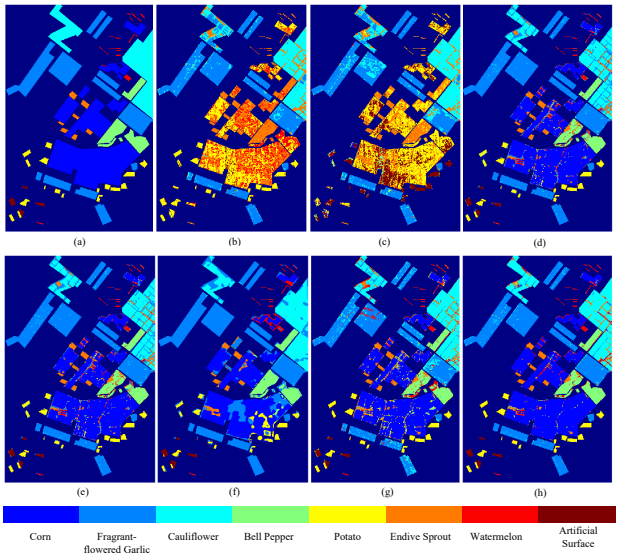


Fig. 9. Visualization of classification results by different methods on Heihe dataset. (a) Groundtruth of Heihe dataset. (b) NPE. (c) LPP. (d) LRE. (e) RPL. (f) S^3PCA. (g) GRSC. (h) LatLRE.

projection techniques, fall short of achieving optimal effectiveness due to their singular focus on individual direction. This limitation arises from the absence of joint consideration for multiple directions. In contrast, the proposed LatLRE takes into account both row and column information of the matrix, which can ensure robustness of the low-rank feature space.

3) According to Eq.(3), it can be seen that although LRE has achieved the exploration of sample relationships, it lacks the excavation of salient features, i.e., LRE ignores the capture of feature correlations. Instead, our LatLRE additionally introduces latent low-rank learning, which takes into account information from both directions of the data. Moreover, from Tables II-IV, one can see that LatLRE outperforms the comparative methods, which shows the effectiveness of mining feature correlations to improve classification performance.

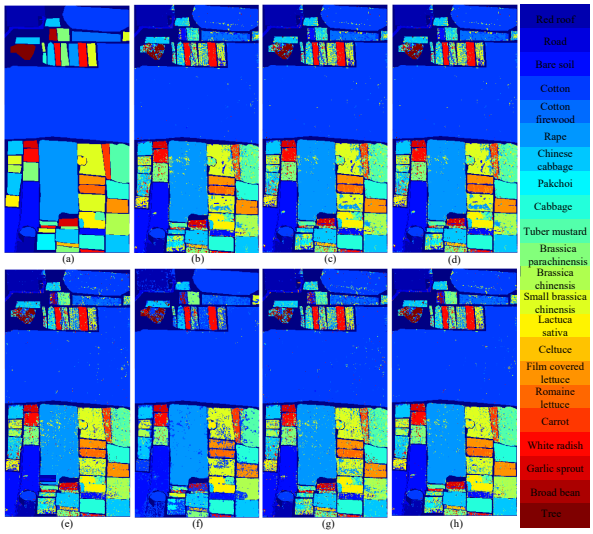


Fig. 10. Visualization of classification results by different methods on WHU-Hi-HongHu dataset. (a) Groundtruth of WHU-Hi-HongHu dataset. (b) NPE. (c) LPP. (d) LRE. (e) RPL. (f) S^3PCA. (g) GRSC. (h) LatLRE.

G. Discussions of Nuclear Norm and Logarithmic Norm

In the proposed method, we introduce the nuclear norm to characterize the low-rankness of the sample relationships matrix $\mathbf{Z}$ and apply the logarithmic norm to describe the low-rankness of the feature correlations matrix $\mathbf{L}$. The reasons are as follows: According to Fig. 2, we observe that the singular value curves of the sample relationships matrix $\mathbf{Z}$ and the feature correlations matrix $\mathbf{L}$ drop rapidly and approach 0, which indicates that they are both low-rank matrices. Moreover, the main singular values of the sample relationship matrix $\mathbf{Z}$ are mainly around 1 and less than 1, while the singular values of the feature correlations matrix $\mathbf{L}$ are relatively large. Therefore, different low-rank functions need to be introduced to accurately characterize the low-rankness of $\mathbf{Z}$ and $\mathbf{L}$, respectively.

Inspired by the rank approximation of different functions in Fig. 3, one can be seen that when the singular values are small, the nuclear function is closer to the rank, while the logarithmic function is closer to the rank when the singular value is relatively large. Therefore, based on the above observation, we introduce the nuclear norm on $\mathbf{Z}$ and the logarithmic norm on $\mathbf{L}$ to accurately characterize their low-rankness.

To test the impact of other potential approaches on the classification results, we conduct comparative experiments on three datasets, the experimental results are shown in Tables V-VII. Specifically, the experimental settings are as follows:

- Apply the nuclear norm on both $\mathbf{Z}$ and $\mathbf{L}$.
- Apply the logarithmic norm on both $\mathbf{Z}$ and $\mathbf{L}$.
- Apply the logarithmic norm on $\mathbf{Z}$ and nuclear norm on $\mathbf{L}$.
- Apply the nuclear norm on $\mathbf{Z}$ and logarithmic norm on $\mathbf{L}$.

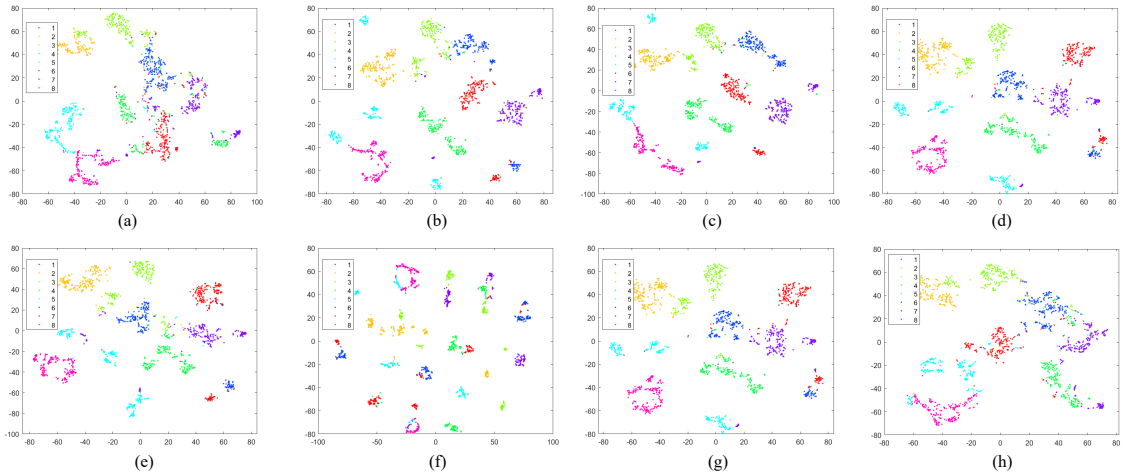


Fig. 11. The t-SNE visualization of the feature matrix after different FDR methods on the Heihe dataset. (a) Original, (b) NPE, (c) LPP, (d) LRE, (e) RPL, (f) S^3PCA, (g) GRSC, (h) LatLRE.

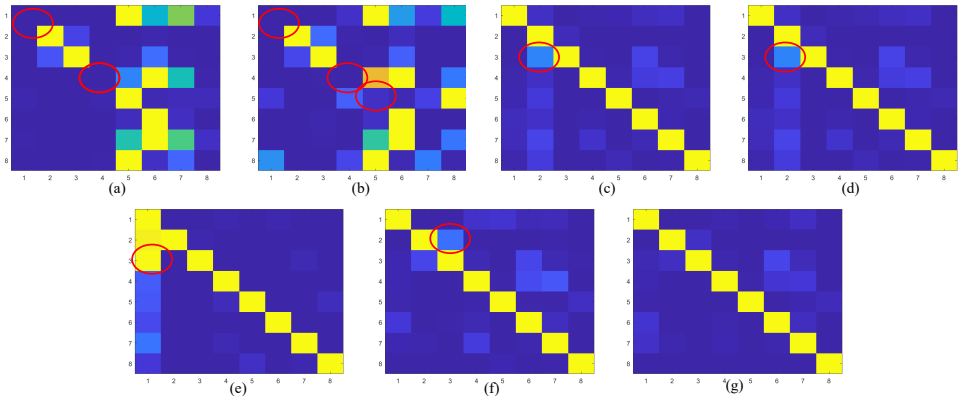


Fig. 12. Visualization of confusion matrix by using different FDR methods on the Heihe dataset. (a) NPE, (b) LPP, (c) LRE, (d) RPL, (e) S^3PCA, (f) GRSC, (g) LatLRE.

Tables V-VII show that the design of $||\mathbf{Z}||_* + ||\mathbf{L}||_L$ achieves the best results in most cases. Meanwhile, the combination of $||\mathbf{Z}||_L + ||\mathbf{L}||_L$ have yielded competitive results.

TABLE V
CLASSIFICATION PERFORMANCE APPLYING DIFFERENT NORMS ON $\mathbf{Z}$ AND $\mathbf{L}$ FOR INDIAN PINES DATASET.

$\mathbf{Z}$		$\mathbf{L}$		OA	AA	κ
$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$	$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$			
✓		✓		95.58	86.04	94.95
	✓		✓	95.36	86.38	94.70
	✓	✓		95.93	87.62	95.36
✓			✓	**96.48**	**87.89**	**95.98**

TABLE VI
CLASSIFICATION PERFORMANCE OF APPLYING DIFFERENT NORMS ON $\mathbf{Z}$ AND $\mathbf{L}$ FOR HEIHE DATASET.

$\mathbf{Z}$		$\mathbf{L}$		OA	AA	κ
$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$	$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$			
✓		✓		88.41	89.77	84.54
	✓		✓	87.50	87.62	83.24
	✓	✓		88.90	89.61	85.22
✓			✓	**91.40**	**92.09**	**88.40**

TABLE VII
CLASSIFICATION PERFORMANCE OF APPLYING DIFFERENT NORMS ON $\mathbf{Z}$ AND $\mathbf{L}$ FOR WHU-HI-HONGHU DATASET.

$\mathbf{Z}$		$\mathbf{L}$		OA	AA	κ
$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$	$\|\|\cdot\|\|_*$	$\|\|\cdot\|\|_L$			
✓		✓		90.09	76.63	87.41
	✓		✓	87.03	68.73	83.46
	✓	✓		89.82	76.26	87.05
✓			✓	**90.11**	**76.69**	**87.44**

H. Analysis of Visualization

To intuitively compare the effectiveness of the propoesd LatLRE, t-SNE [59] has been introduced to visualize the features projected by the comparative methods. As a typical example, Fig. 11 illustrates the visualization results on the Heihe dataset. Samples from the same category are represented by the same color. It can be easily observed that most of the comparative methods exhibit either excessive intra-class variance or overlapping clusters. Specifically, while method S^3PCA ensures inter-class discriminability, it overlooks the importance of intra-class compactness. This can be observed in Fig. 11(f), where the visualization of features after dimensionality reduction using method S^3PCA reveals instances of samples from the same class being divided into different

subclusters. Although LPP and LRE methods show some improvement, they still exhibit similar shortcomings for the fifth class. In contrast, NPE demonstrates overlapping clusters. In contrast to the aforementioned methods, the LatLRE ensures the compactness of intra-class samples. Moreover, we visualize the confusion matrices of different dimensionality reduction methods after classification by an SVM classifier, as shown in Fig. 12. It is noticeable that, except for S^3PCA, other methods have misclassified the second class, while S^3PCA exhibits poor classification performance for the first and third classes. In comparison to these methods, the proposed method LatLRE demonstrates promising result, as shown in Fig. 12(g), where the diagonal structure of the confusion matrix is notably clear. This is because the LatLRE can fully explore the salient features and employ multi-directional projection learning to effectively preserve the data structure. Additionally, the introduction of logarithmic norm helps reduce noise and redundant information.

I. Ablation Study

To explore the effectiveness of each component of the proposed LatLRE, we conducted ablation study. The settings are as follows:

- Case 1: To explore the influence of the projection matrix $\mathbf{P}$ in the LatLRE, we set $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}$ to $\mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E}$.
- Case 2: To explore the influence of the $l_{2,1}$-norm in the LatLRE, we remove the sparsity constraint term $\|\mathbf{E}\|_{2,1}$ and set $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}$ to $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX})$.
- Case 3: To examine the effect of the logarithmic norm in LatLRE, we remove the feature correlations description item $\|\mathbf{L}\|_L$ and set $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}$ to $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{XZ} + \mathbf{E}$.
- Case 4: To explore the impact of the nuclear norm in LatLRE, we remove the sample relationships characterization item $\|\mathbf{Z}\|_*$ and set $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T(\mathbf{XZ} + \mathbf{LX}) + \mathbf{E}$ to $\mathbf{P}^T\mathbf{X} = \mathbf{P}^T\mathbf{LZ} + \mathbf{E}$.
- Case 5: The proposed LatLRE.

TABLE VIII
ABLATION STUDY OF LATLRE ON HEIHE DATASET.

Case	$\mathbf{P}$	$\mathbf{E}$ ($\|\|\mathbf{E}\|\|_{2,1}$)	$\mathbf{L}$ ($\|\|\mathbf{L}\|\|_L$)	$\mathbf{Z}$ ($\|\|\mathbf{Z}\|\|_*$)	OA	AA	κ
1		✓	✓	✓	88.65	86.88	84.69
2	✓		✓	✓	90.15	90.52	86.79
3	✓	✓		✓	88.88	88.97	85.15
4	✓	✓	✓		90.45	90.71	87.16
5	✓	✓	✓	✓	**91.40**	**92.09**	**88.40**

Taking the Heihe dataset as an example, the results of the ablation experiments are shown in Table VIII. The comparison between Case 1 and Case 5 demonstrates the effect of the projection matrix $\mathbf{P}$, and it can be seen that the addition of projection learning has led to an improvement in the experimental results, with OA, AA, and κ improving by 2.75%, 5.21%, and 3.71%, respectively. This improvement occurs because projection learning effectively alleviates the dimensionality catastrophe problem in hyperspectral images. Furthermore, comparing Case 2 with Case 5 reveals the effectiveness of $||\mathbf{E}||_{2,1}$ in suppressing noise interference and improving model robustness. The negative impacts of removing the feature correlations matrix $\mathbf{L}$ and sample relationships matrix $\mathbf{Z}$ are shown in Case 3 and Case 4. It is obvious that capturing both feature correlations and sample relations can enhance the classification performance, where the overall accuracy is decreased by 2.16% and 0.59%, respectively. Overall, the experiments above demonstrate that the projection matrix $\mathbf{P}$, the sparsity constraint term $||\mathbf{E}||_{2,1}$, the feature correlations description item $||\mathbf{L}||_L$, and the sample relationships description item $||\mathbf{Z}||_*$ are crucial for the effectiveness of LatLRE.

V. CONCLUSION

In this paper, a representation-based method named LatLRE is proposed for robust FDR, which can exploit the intrinsic subspace structure of data to enhance the robustness. In particular, the proposed LatLRE model can learn an optimal projection matrix to map the data into a low-dimensional space with the noise filtered. By using the nuclear norm and Logarithmic norm to substitute the two underlying rank functions respectively, the optimization of the LatLRE model becomes more reasonable and more flexible for singular values, which also improves the efficiency of both global structure and feature correlations. Experimental results show that the proposed method can obtain better classification accuracy than the widely-used FDR methods considered in this paper.

In the future, we will extend LatLRE into the tensor space to directly deal with the high-dimensional structure of HSIs and exploit the unified framework of FDR and classification tasks.

REFERENCES

[1] S. P. Ang, S. L. Phung, L. Bui, and A. Bouzerdoum, "Adaptornas: A new perturbation-based neural architecture search for hyperspectral image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 3, pp. 1559–1571, Mar. 2024.

[2] B. Yun, B. Lei, J. Chen, H. Wang, S. Qiu, W. Shen, Q. Li, and Y. Wang, "SpecTr: Spectral transformer for microscopic hyperspectral pathology image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 6, pp. 4610–4624, Jun. 2024.

[3] H. Chen, W. Zhao, T. Xu, G. Shi, S. Zhou, P. Liu, and J. Li, "Spectral-wise implicit neural representation for hyperspectral image reconstruction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 3714–3727, May 2024.

[4] Y. Zhang, Y. Wang, X. Chen, X. Jiang, and Y. Zhou, "Spectral-spatial feature extraction with dual graph autoencoder for hyperspectral image clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8500–8511, Dec. 2022.

[5] G. Jiang, Y. Zhang, X. Wang, X. Jiang, and L. Zhang, "Structured anchor learning for large-scale hyperspectral image projected clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 3, pp. 2328–2340, Mar. 2025.

[6] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 641–656, 2021.

[7] J.-Y. Yang, H.-C. Li, J.-H. Yang, L. Pan, Q. Du, and A. Plaza, "Multifrequency graph convolutional network with cross-modality mutual enhancement for multisource remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024.

[8] Y. Duan, C. Chen, M. Fu, Y. Li, X. Gong, and F. Luo, "Dimensionality reduction via multiple neighborhood-aware nonlinear collaborative analysis for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 10, pp. 9356–9370, Oct. 2024.

[9] Y.-J. Deng, M.-L. Yang, H.-C. Li, C.-F. Long, K. Fang, and Q. Du, "Feature dimensionality reduction with $l_{2,p}$-norm-based robust embedding regression for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, no. 5509314, 2024.

[10] W. Dong, T. Yang, J. Qu, T. Zhang, S. Xiao, and Y. Li, "Joint contextual representation model-informed interpretable network with dictionary aligning for hyperspectral and LiDAR classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 11, pp. 6804–6818, Nov. 2023.

[11] J. Fan, T. Chen, and S. Lu, "Superpixel guided deep-sparse-representation learning for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3163–3173, Nov. 2018.

[12] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 19, no. 10, pp. 1768–1782, Oct. 2008.

[13] T. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[14] H. Zhao, Z. Wang, and F. Nie, "A new formulation of linear discriminant analysis for robust dimensionality reduction," *IEEE Trans. Know. Eng.*, vol. 31, no. 4, pp. 629–640, Apr. 2019.

[15] J. Wen, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.

[16] C.-N. Li, Y. H. Shao, W. Yin, and M.-Z. Liu, "Robust and sparse linear discriminant analysis via an alternating direction method of multipliers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 915–926, Mar. 2020.

[17] Y. Li, J. Qu, W. Dong, and Y. Zheng, "Hyperspectral pansharpening via improved PCA approach and optimal weighted fusion strategy," *Neurocomput.*, vol. 315, pp. 371–380, Nov. 2018.

[18] N. Kwak, "Principal component analysis based on l_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[19] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust pca in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.

[20] N. Kwak, "Principal component analysis by l_p-norm maximization," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 594–609, May 2014.

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

[22] Y. Lu, C. Yuan, Z. Lai, X. Li, W. Wong, and D. Zhang, "Nuclear norm-based 2DLPP for image classification," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2391–2403, Nov. 2017.

[23] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "Superpca: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[24] X. Zhang, X. Jiang, J. Jiang, Y. Zhang, X. Liu, and Z. Cai, "Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 5502210, Aug. 2022.

[25] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2004, pp. 153–160.

[26] G. Shi, H. Huang, and L. Wang, "Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1425–1429, Aug. 2019.

[27] J. Wang, F. Xie, F. Nie, and X. Li, "Unsupervised adaptive embedding for dimensionality reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6844–6855, Nov. 2021.

[28] J. Wen, Z. Tian, H. She, and W. Yan, "Feature extraction of hyperspectral images based on preserving neighborhood discriminant embedding," in *Proc. IEEE IASP*, Apr. 2010, pp. 257–262.

[29] Z. Zhang, F. Li, M. Zhao, L. Zhang, and S. Yan, "Robust neighborhood preserving projection by nuclear/$l_{2,1}$-norm regularization for image feature extraction," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1607–1622, Apr. 2017.

[30] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2143–2156, Dec. 2007.

[31] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[32] X. Jiang, L. Xiong, W. Yan, Y. Zhang, X. Liu, and Z. Ca, "Unsupervised dimensionality reduction for hyperspectral imagery via laplacian regularized collaborative representation projection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[33] K. Chen, G. Yang, J. Wang, Q. Du, and H. Su, "Unsupervised dimensionality reduction with multifeature structure joint preserving embedding for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens*, vol. 16, pp. 7585–7599, 2023.

[34] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[35] Z. Ren, Q. Sun, B. Wu, X. Zhang, and W. Yan, "Learning latent low-rank and sparse embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 2094–2107, Sep. 2020.

[36] N. Ly, Q. Du, and J. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jul. 2014.

[37] W. He, H. Zhang, L. Zhang, W. Philips, and W. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, Mar. 2016.

[38] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[39] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[40] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," *Pattern Recogn.*, vol. 81, pp. 326–340, Sep. 2018.

[41] J. Lu, H. Wang, J. Zhou, Y. Chen, Z. Lai, and Q. Hu, "Low-rank adaptive graph embedding for unsupervised feature extraction," *Pattern Recogn.*, vol. 113, no. 107758, May 2021.

[42] W. Wong, Z. Lai, J. Wen, X. Fang, and Y. Lu, "Low-rank embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2905–2917, Jun. 2017.

[43] L. Xie, M. Yin, X. Yin, Y. Liu, and G. Yin, "Low-rank sparse preserving projections for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5261–5274, Nov. 2018.

[44] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.

[45] S. Yu and W. Yiquan, "Subspace clustering based on latent low rank representation with frobenius norm minimization," *Neurocomput.*, vol. 275, pp. 2479–2489, Jan. 2018.

[46] Z. Liu, D. Hu, Z. Wang, J. Gou, and T. Jia, "LatLRR for subspace clustering via reweighted frobenius norm minimization," *Expert Syst. Appl.*, vol. 224, no. 119977, Aug. 2023.

[47] Q. Shen, Y. Chen, Y. Liang, S. Yi, and W. Liu, "Weighted schatten p-norm minimization with logarithmic constraint for subspace clustering," *Signal Process.*, vol. 198, no. 108568, Apr. 2022.

[48] M. Wang, J. Yu, L. Niu, and W. Sun, "Feature extraction for hyperspectral images using low-rank representation with neighborhood preserving regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 836–840, Jun. 2017.

[49] L. Chen, X. Jiang, X. Liu, and Z. Zhou, "Logarithmic norm regularized low-rank factorization for matrix and tensor completion," *IEEE Trans. Image Process.*, vol. 30, pp. 3434–3449, Mar. 2021.

[50] L. Yang, J. Miao, and K. Kou, "Quaternion-based color image completion via logarithmic approximation," *Inf. Sci.*, vol. 588, pp. 82–105, Sep. 2022.

[51] J.-H. Yang, C. Chen, H. Dai, M. Ding, Z. Wu, and Z. Zheng, "Robust corrupted data recovery and clustering via generalized transformed tensor low-rank representation," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 35, no. 7, pp. 8839–8853, 2024.

[52] J.-H. Yang, Y. Zhou, L. F. Zhang, and H.-C. Li, "Mixed-noise robust tensor multi-view clustering via adaptive dictionary learning," *Inform. Fusion*, vol. 123, pp. 1–16, 2025.

[53] X.-L. Zhao, J.-H. Yang, T.-H. Ma, T.-X. Jiang, and M. Ng, "Tensor completion via complementary global, local, and nonlocal priors," *IEEE Trans. Image Process.*, vol. 31, pp. 984–999, 2022.

[54] Z. Lin and M. C. Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Tech. Rep.*, Mar. 2009.

[55] D. G. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1997.

[56] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 2005, pp. 1208–1213.

[57] X. Song, H.-C. Li, L. Pan, Y.-J. Deng, P. Zhang, L. You, and Q. Du, "Unsupervised robust projection learning by low-rank and sparse decomposition for hyperspectral feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, May 2021.
[58] J. Wang, C. Tang, X. Zheng, X. Liu, W. Zhang, and E. Zhu, "Graph regularized spatial–spectral subspace clustering for hyperspectral band selection," *Neural Netw.*, vol. 153, pp. 292–302, Sep. 2022.
[59] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

Heng-Chao Li (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently a Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. His research interests include statistical analysis of synthetic aperture radar (SAR) images, remote sensing image processing, and pattern recognition.

Jun-Qiu Wang is currently pursuing the Ph.D. degree with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. His research interests include hyperspectral image processing, feature selection, and data mining.

Si-Jia Xiang is currently pursuing the Ph.D. degree with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. Her research interests include hyperspectral image classification, multi-view clustering, and data mining.

Jing-Hua Yang received the Ph.D. degree from the Macau University of Science and Technology, Macau, China, in 2023. She is currently an assistant professor with the School of Information Science and Technology, Southwest Jiaotong University. Her current research interests include data mining, image processing, and deep learning.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000. She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.