SIAM J. Imaging Sciences
Vol. 17, No. 2, pp. 917–950

Generalized Nonconvex Hyperspectral Anomaly Detection via Background Representation Learning with Dictionary Constraint*

Quan Yu† and Minru Bai‡

Abstract. Anomaly detection in the hyperspectral images, which aims to separate interesting sparse anomalies from backgrounds, is a significant topic in remote sensing. In this paper, we propose a generalized nonconvex background representation learning with dictionary constraint (GNBRL) model for hyperspectral anomaly detection. Unlike existing methods that use a specific nonconvex function for a low rank term, GNBRL uses a class of nonconvex functions for both low rank and sparse terms simultaneously, which can better capture the low rank structure of the background and the sparsity of the anomaly. In addition, GNBRL simultaneously learns the dictionary and anomaly tensor in a unified framework by imposing a three-dimensional correlated total variation constraint on the dictionary tensor to enhance the quality of representation. An extrapolated linearized alternating direction method of multipliers (ELADMM) algorithm is then developed to solve the proposed GNBRL model. Finally, a novel coarse to fine two-stage framework is proposed to enhance the GNBRL model by exploiting the nonlocal similarity of the hyperspectral data. Theoretically, we establish an error bound for the GNBRL model and show that this error bound can be superior to those of similar models based on Tucker rank. We prove that the sequence generated by the proposed ELADMM algorithm converges to a Karush–Kuhn–Tucker point of the GNBRL model. This is a challenging task due to the nonconvexity of the objective function. Experiments on hyperspectral image datasets demonstrate that our proposed method outperforms several state-of-the-art methods in terms of detection accuracy.

Key words. hyperspectral anomaly detection, generalized nonconvex functions, dictionary, coarse to fine two-stage framework, error bound

MSC codes. 15A69, 47A80, 65K05

DOI. 10.1137/23M157363X

1. Introduction.

Hyperspectral images (HSIs) are a powerful tool for detecting and locating ground objects and providing spatial positioning and structural information [3]. They have been successfully applied in a variety of fields, including classification [18, 30, 35], image fusion [6], and target detection [50, 52, 65]. In recent years, hyperspectral anomaly detection (HAD) has attracted considerable interest due to its significance for public safety and defense [19, 20, 58]. The goal of HAD is to identify and separate anomalous objects from the background by exploring the differences between anomalies and their surroundings. To tackle this

*Received by the editors June 20, 2023; accepted for publication (in revised form) January 18, 2024; published electronically April 12, 2024.
https://doi.org/10.1137/23M157363X
Funding: The work of the authors was funded by National Natural Science Foundation of China (NSFC) grants 11971159 and 12071399 and Hunan Provincial Key Laboratory of Intelligent Information Processing and Applied Mathematics.
†School of Mathematics, Hunan University, Changsha, Hunan Province, 410082, China (quanyu@hnu.edu.cn).
‡Corresponding author. School of Mathematics, Hunan University, Changsha, Hunan Province, 410082, China (minru-bai@hnu.edu.cn).

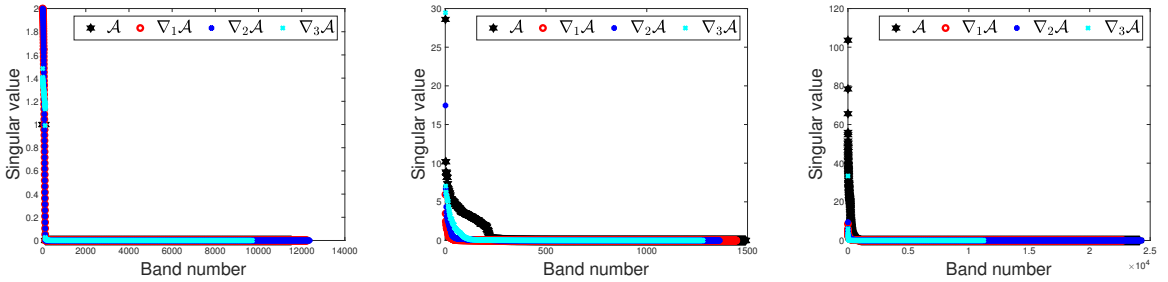
problem, researchers have proposed various methods, such as statistic based methods, deep learning based methods, low rank matrix/tensor decomposition based methods, and low rank matrix/tensor representation based methods.

A common statistic based algorithm is the Reed–Xiaoli (RX) detector [40]. It calculates the mean vector and covariance matrix of the samples from the HSI to obtain the Mahalanobis distance between testing pixels and their surrounding background. One limitation of RX is that it relies on the assumption that the background can be characterized by a single normal distribution, which is seldom the case for real hyperspectral applications [20]. Therefore, various modified RX methods have been proposed, such as weighted-RX [15], kernel-RX [27], segmented-RX [33], and subspace-RX [42]. However, these methods rely on manually designed distribution forms, which may not capture the complex and diverse background characteristics. To overcome this limitation, deep learning based methods have been adopted for HAD tasks, which use deep networks to mine and interpret higher order information contained in the HSI. Deep learning based methods can be categorized into unsupervised, supervised, or self-supervised learning. Unsupervised methods learn background representations or anomalies using autoencoder [63] or generative adversarial networks [21]. Supervised methods use a convolutional neural network to extract features and classify anomalies using labeled data [29]. Self-supervised methods generate pseudolabels by predicting some tasks without labeled data. For example, BS^{3}LNet [13] and PDBSNet [47] predicted blind-spot pixels in HSI and used them to train the model.

To separate the background and anomaly, some researchers have used low rank matrix decomposition based methods that have shown good potential. Sun et al. [45] and Zhang et al. [62] used the robust principal component analysis (RPCA) framework to decompose the HSI data into low rank and sparse parts. But they still failed to effectively distinguish weak anomalies from noise. Recently, matrix representation based methods have been successfully applied in HAD. These methods construct a fixed background dictionary using different techniques, and then formulate the target detection model using this dictionary. For example, Xu et al. [53] and Cheng and Wang [9] obtained the background dictionary by applying the k-means clustering method, while Zhuang et al. [65] used the singular value decomposition (SVD) method to get it.

HSIs are inherently considered as third order tensors, where the height and width modes represent the spatial characteristics and the spectral mode represents the spectral characteristics. Compared with matrix based HAD methods, tensor based HAD methods treat the original HSI data as a whole, and preserve the spatial structure. In recent years, tensor based methods have attracted more attention in HAD. Chen, Yang, and Wang [8] and Li et al. [28] proposed a tensor RPCA based method to separate a principal component part and a residual part based on tensor decomposition. Wang et al. [49] developed a tensor representation based method, which first constructed a fixed background dictionary using the tensor RPCA method, and then formulated the target detection model using this dictionary.

In representation based methods, the dictionary plays a critical role. There are three main methods of dictionary construction in current research: SVD [65], k-means clustering [9, 53], and (tensor) RPCA [49]. A drawback of these methods is that they depend on a prebuilt dictionary that compromises the quality of representation. Therefore, we need to explore the features of the dictionary so that we can perform dictionary construction and anomaly



(a) SVD [65] (b) k-means clustering [9, 53] (c) RPCA [49]

Figure 1. *Singular values of dictionary tensor $\mathcal{A}$ and gradient tensor $\nabla_u\mathcal{A}$, $u \in [3]$. The definition of $\nabla_u\mathcal{A}$ is given in (3.3).*

detection simultaneously. In this way, we can update the dictionary in real time at each step to achieve the best results. With the development of low rank recovery theory, Peng et al. [37] used the three-dimensional correlated total variation (3DCTV) regularization to exploit the structural characteristics of the HSI data. Based on t-SVD [25], we show the singular values of the dictionary tensor and those of the gradient tensor obtained by a differential operator in Figure 1. From Figure 1, we can see that both the dictionary tensors generated by k-means clustering and (tensor) RPCA have a low rank structure, and gradient tensors are more low rank than the original dictionary tensor. Therefore, we use the low rank property of gradient tensors to describe the dictionary tensor.

One common challenge is that almost all of matrix/tensor based HAD methods use the nuclear norm and l_1-norm to investigate the low rank and sparse characteristic. For example, the studies in [9, 53] exploited the convex matrix nuclear norm and l_1-norm while those of [44, 48] used the convex tensor nuclear norm and l_1-norm. However, the nuclear norm is essentially the l_1-norm of all singular values, which is known to yield biased estimators and cannot achieve the best estimation performance. To address this issue, inspired by the significant performance of nonconvex regularization in vector [1, 36], matrix [56], and tensor problems [55, 64], in this paper, we propose a generalized nonconvex tensor approximation of the tensor nuclear norm (TNN) [43] and l_1-norm. Unlike most existing nonconvex HAD methods [28, 49], which only use nonconvex approximation on low rank, we apply nonconvex approximation to both low rank and sparsity. In addition, most existing nonconvex HAD methods were solved case-by-case. We also provide a general solver with a convergence guarantee under mild conditions.

Such two concerns are not well solved in existing tensor representation based methods. In this paper, we propose a novel generalized nonconvex background representation learning with dictionary constraint (GNBRL) model. GNBRL assumes that the background part is global low rank. However, this may lead to a trade-off between restoring the original low rank parts and neglecting the potentially high rank parts. To overcome this limitation, we develop a novel coarse to fine two-stage (CF2) framework for GNBRL. In the first stage, GNBRL is applied to the whole HSI to obtain a coarse result and detect most anomalies. In the second stage, the HSI is divided into smaller patches based on block matching three dimensional (BM3D) [10, 11] and each patch is solved independently using GNBRL. The restored patches then

replace their counterparts in the coarse result if they satisfy a certain metric. We emphasize that this divide-and-conquer strategy can improve the performance of existing HAD methods.

Compared with existing matrix and tensor based methods, the main contributions of this study are summarized below.

(1) We propose a GNBRL model that simultaneously learns the dictionary and anomaly tensor in a unified framework, which can enhance the quality of representation. By employing a class of generalized nonconvex functions as the TNN and l_1-norm approximations, we can capture the low rank structure of the background and the sparsity of the anomaly more accurately.

(2) We provide an error bound of the anomaly tensor recovered by the GNBRL model. Moreover, we develop an extrapolated linearized alternating direction method of multipliers (ELADMM) algorithm to solve the GNBRL model, and the convergence analysis is also given. This is a challenging task due to the nonconvexity of the objective function. These theoretical results are rarely given by previous HAD works.

(3) We propose a CF2 framework for GNBRL, which can greatly improve its performance by first obtaining the coarse anomaly tensor and then refining the anomaly tensor.

The remainder of this paper is organized as follows. First, in section 2, some notations and preliminaries of tensors are introduced. Next, in section 3, a GNBRL model is presented and its worst-case error bound is established. In addition, we propose a CF2 framework to improve the proposed GNBRL model in this section. After that, in section 4, an ELADMM algorithm is proposed to solve the GNBRL model and give the convergence analysis. Subsequently, in section 5, extensive numerical experiments are reported to verify the superior performance of the GNBRL and CF2-GNBRL model. Finally, the concluding remarks are given in section 6.

2. Preliminary knowledge on tensor. Before proceeding, we first present some notations here. For a positive integer n, $[n] := \{1, 2, \ldots, n\}$. Scalars, vectors, and matrices are denoted as lowercase letters (a), boldface lowercase letters ($\boldsymbol{a}$), and uppercase letters (A), respectively. Third order tensors are denoted as calligraphic letters ($\mathcal{A}$), and the set of all the third order real tensors and complex tensors are denoted as $\mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathbb{C}^{n_1\times n_2\times n_3}$, respectively. For a third order tensor $\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$, we use the notations $X^{(k)}$, $k\in[n_3]$ to denote its kth frontal slice. To avoid confusion, we use $\mathcal{X}_{ijk}$ and $\mathcal{X}(i,j,k)$ to denote the (i,j,k)th entries of $\mathcal{X}$. The inner product of two tensors $\mathcal{X},\mathcal{Y}\in\mathbb{R}^{n_1\times n_2\times n_3}$ is the sum of products of their entries, i.e., $\langle\mathcal{X},\mathcal{Y}\rangle=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{X}_{ijk}\mathcal{Y}_{ijk}$. We use $\bar{\mathcal{X}}$ to denote the discrete Fourier transformation (DFT) along the third dimension of $\mathcal{X}$, which can be obtained by MATLAB command "fft", that is, $\bar{\mathcal{X}}=\mathrm{fft}(\mathcal{X},[\,],3)$, and the inverse operation is $\mathcal{X}=\mathrm{ifft}(\bar{\mathcal{X}},[\,],3)$.

By defining the block circulant matrix $bcirc(\mathcal{X})$ as

$$bcirc(\mathcal{X})=\begin{bmatrix} X^{(1)} & X^{(n_3)} & \cdots & X^{(2)} \\ X^{(2)} & X^{(1)} & \cdots & X^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ X^{(n_3)} & X^{(n_3-1)} & \cdots & X^{(1)} \end{bmatrix},$$

we can then introduce the tensor spectral norm and the t-product between two third order tensors.

Definition 2.1 (tensor spectral norm [31]). *The tensor spectral norm of* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is defined as* $\|\mathcal{X}\| = \|bcirc(\mathcal{X})\|$.

Definition 2.2 (t-product [25]). *The t-product between* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *and* $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ *is defined as*

$$\mathcal{X} * \mathcal{Y} = fold\,(bcirc(\mathcal{X}) \cdot unfold\,(\mathcal{Y})) \in \mathbb{R}^{n_1 \times n_4 \times n_3},$$

where $unfold(\mathcal{Y}) = [Y^{(1)}; Y^{(2)}; \ldots; Y^{(n_3)}] \in \mathbb{R}^{n_2 n_3 \times n_4}$ *and its inverse operator fold is defined as* $fold(unfold(\mathcal{Y})) = \mathcal{Y}$.

Based on the definition of the t-product, we will further introduce a new tensor decomposition framework, t-SVD.

Definition 2.3 (t-SVD [25]). *For any given tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *it can be factorized as*

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H,$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ *are orthogonal tensors and* $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is an f-diagonal tensor.*

Some related concepts, such as the f-diagonal tensor, the conjugate transpose, and so on, are explained in Appendix A.

Definition 2.4 (TNN [43]). *The TNN of a tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is the sum of singular values of all front slices of* $\bar{\mathcal{X}}$, *that is,* $\|\mathcal{X}\|_{TNN} = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^{\min\{n_1, n_2\}} \sigma_i(\bar{X}^{(k)})$.

3. A GNBRL model for HAD.

In this section, we first present a GNBRL model for HAD, and then propose a CF2 framework for GNBRL. Finally, an error bound analysis of GNBRL is given.

3.1. Generalized nonconvex model with dictionary constraint.

Recent advances have shown that tensor representation based methods for HAD outperform matrix representation based methods. This is because HSI data are inherently a third order tensor, and matrix representation based methods destroy the intrinsic HSI tensor structure. The seminal works in this area are [17] and [49], both of which follow the general model

$$\min_{\mathcal{L}, \mathcal{S}} \operatorname{rank}(\mathcal{L}) + \lambda \|\mathcal{S}\|_{\ell_{F,0}} \quad \text{s.t.} \quad \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}, \tag{3.1}$$

where $\mathcal{A} * \mathcal{L}$ is the background part, $\mathcal{A}$ denotes the background dictionary, $\mathcal{L}$ is the corresponding representation coefficients, and S represents the sparse part related to anomalies. The tensor group sparsity $\ell_{F,0}$-norm is defined as $\|\mathcal{X}\|_{\ell_{F,0}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|\mathcal{X}(i, j, :)\|_F^0$ (adopting the convenience that $0^0 = 0$).

Differently from model (3.1) which simply adopts a prebuilt dictionary $\mathcal{A}$, we perform dictionary construction and anomaly detection simultaneously to learn a more comprehensive dictionary for background reconstruction. The low rank structure of the dictionary tensors is evident from Figure 1 in section 1. The gradient tensors, which capture the variations of the dictionary tensors, have an even lower rank. We exploit this property to model the dictionary tensors using the gradient tensors, and it can be modeled as follows,

$$\min_{\mathcal{A},\mathcal{L},\mathcal{S}} \sum_{u=1}^{3} \alpha_u \text{rank}\left(\nabla_u \mathcal{A}\right) + \lambda_1 \text{rank}\left(\mathcal{L}\right) + \lambda_2 \left\|\mathcal{S}\right\|_{\ell_{F,0}} \quad \text{s.t.} \quad \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}, \tag{3.2}$$

where $\nabla_1\mathcal{A}$, $\nabla_2\mathcal{A}$, $\nabla_3\mathcal{A}$ denote the first order forward finite-difference operators along the vertical, horizontal, and spectral directions, respectively. Here, for the (i,j,k)th entry, each forward finite-difference operator is defined by

$$\begin{cases} \nabla_1\mathcal{A}(i,:,:) = \mathcal{A}(i+1,:,:) - \mathcal{A}(i,:,:), & i \in [n_1 - 1], \\ \nabla_2\mathcal{A}(:,j,:) = \mathcal{A}(:,j+1,:) - \mathcal{A}(:,j,:), & j \in [n_2 - 1], \\ \nabla_3\mathcal{A}(:,:,k) = \mathcal{A}(:,:,k+1) - \mathcal{A}(:,:,k), & k \in [n_3 - 1]. \end{cases} \tag{3.3}$$

Here, we use zero paddings for $\mathcal{A}$ before applying the differential operation on it, i.e., $\nabla_1\mathcal{A}(n_1,:,:) = 0$, $\nabla_2\mathcal{A}(:,n_2,:) = 0$, $\nabla_3\mathcal{A}(:,:,n_3) = 0$, which keep the size of $\nabla_u\mathcal{A}$ the same as $\mathcal{A}$ and makes calculation convenient in the subsequent operations.

Differently from most existing nonconvex HAD methods [17, 28, 49] that only use nonconvex approximation for low rank, we apply it to both low rank and sparsity. Moreover, we pursue a general surrogate for approximating low rank and sparsity, i.e.,

$$\begin{aligned} \left\|\mathcal{X}\right\|_{\psi} &= \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^{\min\{n_1,n_2\}} \psi\left(\sigma_i\left(\bar{X}^{(k)}\right)\right), \\ \left\|\mathcal{X}\right\|_{\ell_{F,1}^{\psi}} &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \psi\left(\left\|\mathcal{X}(i,j,:)\right\|_F\right) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \psi\left(\left(\sum_{k=1}^{n_3} \mathcal{X}(i,j,k)^2\right)^{1/2}\right), \end{aligned} \tag{3.4}$$

where $\psi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a function. This allows us to better capture the low rank structure of the background and the sparsity of the anomaly. Indeed, when $\psi(x) = x$, $\left\|\mathcal{X}\right\|_{\psi}$ and $\left\|\mathcal{X}\right\|_{\ell_{F,1}^{\psi}}$ would degrade into $\left\|\mathcal{X}\right\|_{TNN}$ and $\left\|\mathcal{X}\right\|_{\ell_{F,1}}$, respectively. By replacing the rank$(\cdot)$ and $\|\cdot\|_{\ell_{F,0}}$ with (3.4), our model is formulated as

$$\min_{\mathcal{A},\mathcal{L},\mathcal{S}} \sum_{u=1}^{3} \alpha_u \left\|\nabla_u\mathcal{A}\right\|_{\psi} + \lambda_1 \left\|\mathcal{L}\right\|_{\psi} + \lambda_2 \left\|\mathcal{S}\right\|_{\ell_{F,1}^{\psi}} \quad \text{s.t.} \quad \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}. \tag{3.5}$$

If there is no special explanation, we suppose that Assumption 3.1 holds throughout the paper.

Assumption 3.1. The function $\psi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies: ψ is continuous, nondecreasing, and concave with $\psi(0) = 0$.

Theorem 3.2. *Most functions satisfy Assumption* 3.1. *We list five of them as follows. Here we only consider the case* $x > 0$.

(1) *L1:* $\psi^{L1}(x) = x$ [46];
(2) *Lp:* $\psi^{\text{Lp}}(x) = x^p$, $p \in (0,1)$ [7];
(3) *MCP:* $\psi^{\text{MCP}}(x) = \begin{cases} x - \frac{x^2}{2\alpha}, & 0 \le x \le \alpha, \\ \frac{\alpha}{2}, & x > \alpha \end{cases}$ *with* $\alpha > 0$ [60];
(4) *Logarithm:* $\psi^{\text{Log}}(x) = \log(\frac{x}{\theta} + 1)$ *with* $\theta > 0$ [12, 14];

(5) *Capped folded functions:*
- *Capped L1:* $\psi^{CapL1}(x) = \min\{1, \frac{x}{v}\}$ [1, 36, 56];
- *Capped Lp:* $\psi^{CapLp}(x) = \min\{1, \frac{x^p}{v^p}\}$, $p \in (0,1)$ [36];
- *Capped MCP:* $\psi^{CapMCP}(x) = \min\{1, \frac{2\alpha}{\nu(2\alpha-\nu)}\psi^{\mathrm{MCP}}(x)\}$, $0 < \nu < \alpha$ [36];
- *Capped logarithm:* $\psi^{CapLog}(x) = \min\{1, \frac{1}{\psi^{\mathrm{Log}}(v)}\psi^{\mathrm{Log}}(x)\}$.

The proof of Theorem 3.2 is straightforward, so we omit it here. For the special forms of ψ given by Theorem 3.2, their proximal mappings often have analytical expressions, which we summarize in Appendix B.

3.2. CF2 framework for GNBRL. GNBRL decomposes the HSI into a low rank component that represents the background and a sparse component that represents the anomalies. However, this method assumes that the background has a global low rank structure, which may ignore some complex objects in the scene and reduce the detection accuracy. To remove this limitation, we develop a CF2 framework for GNBRL (CF2-GNBRL). Here we present the details of the CF2 strategy, which consist of a coarse stage and a fine stage.

1) Coarse stage: In the coarse stage, a coarse anomaly $\tilde{\mathcal{S}}$ is obtained by applying the GNBRL model to the whole HSI.
2) Fine stage: We first divide the whole HSI into N patches of third order subtensors according to BM3D [10, 11]. Then we apply the GNBRL model to each subtensor to obtain $\hat{\mathcal{S}}^1_{patch}, \hat{\mathcal{S}}^2_{patch}, \ldots, \hat{\mathcal{S}}^N_{patch}$. Next, we divide $\tilde{\mathcal{S}}$ into N patches following the partitions employed in the current fine stage to obtain $\tilde{\mathcal{S}}^1_{patch}, \tilde{\mathcal{S}}^2_{patch}, \ldots, \tilde{\mathcal{S}}^N_{patch}$. Finally, we obtain $\mathcal{S}^\star$ by

$$\mathcal{S}^{\star,l}_{patch} = \begin{cases} \tilde{\mathcal{S}}^l_{patch} \text{if gap}\left(\tilde{\mathcal{S}}^l_{patch}, \hat{\mathcal{S}}^l_{patch}\right) < \varrho, \\ \hat{\mathcal{S}}^l_{patch} \text{if gap}\left(\tilde{\mathcal{S}}^l_{patch}, \hat{\mathcal{S}}^l_{patch}\right) \geq \varrho, \end{cases} \tag{3.6}$$

where ϱ is a given parameter and

$$\text{gap}\left(\tilde{\mathcal{S}}^l_{patch}, \hat{\mathcal{S}}^l_{patch}\right) = \frac{\left\|\tilde{\mathcal{S}}^l_{patch} - \hat{\mathcal{S}}^l_{patch}\right\|_F}{\left\|\tilde{\mathcal{S}}^l_{patch}\right\|_F}.$$

Remark 3.3. Similarly to [65], we adopt the BM3D method for tensor patch dividing in the fine stage of the CF2 framework, as it is very fast. We do not explore other tensor patch dividing methods that might improve the performance of the proposed CF2 framework, because we prioritize the trade-off between the detection accuracy and speed. Figure 8 shows that the CF2-GNBRL method has a fast numerical convergence speed and its convergence can be numerically guaranteed.

Remark 3.4. Instead of applying the nonlocal prior directly on the background [16], we use (3.6) to balance between the potentially low rank and high rank parts of the background. This scheme is motivated by the fact that the background obtained in the coarse stage reflects the overall structure of the underlying true situation. If the gap between the patch $\hat{\mathcal{S}}^l_{patch}$ obtained by the fine stage and the corresponding patch in $\tilde{\mathcal{S}}$ is small, it means that the corresponding backgrounds have a large gap, and then $\hat{\mathcal{S}}^l_{patch}$ has a high risk of deviating from

the true value. Therefore, we only take $\hat{\mathcal{S}}^l_{patch}$ as the final value when the gap between $\tilde{\mathcal{S}}^l_{patch}$ and $\hat{\mathcal{S}}^l_{patch}$ is large, otherwise we take $\tilde{\mathcal{S}}^l_{patch}$ as the final value.

3.3. Error bound of GNBRL. In this subsection, we establish a worst-case error bound to measure the average errors of the recovered anomaly tensor in the worst case. The concept of sparse eigenvalues of a matrix was introduced by Bickel, Ritov, and Tsybakov [2] and Meinshausen and Yu [34] to study the perfect recovery property. Similarly to the restricted eigenvalue condition (REC) of a matrix [2, 34], we introduce the notion of the ψ-restricted tensor eigenvalue condition (ψ-RTEC).

Definition 3.5. *The tensor $\mathcal{A}$ is said to satisfy the ψ-RTEC relative to s (ψ-RTEC(s)) if*

$$\zeta_{\min}(2s) = \min_{\|\mathcal{L}\|_\psi \le 2s} \frac{\|\mathcal{A} * \mathcal{L}\|_F}{\|\mathcal{L}\|_F} > 0. \tag{3.7}$$

It is well known that the REC is a sufficient condition for the uniqueness of the sparse solution of the linear system $A\boldsymbol{x} = \boldsymbol{b}$. In order to improve the understanding of the unique properties of the structural low rank solutions, we introduce the concept of the uniquely solvable-type properties of the tensor $\mathcal{A}$.

Definition 3.6. *The tensor $\mathcal{A}$ is said to satisfy the uniquely solvable property relative to s ψ-low rank (ψ-USP(s)) if the linear inverse problem $\mathcal{A} * \mathcal{L} = \mathcal{B}$ has a unique ψ-USP solution satisfying $\|\mathcal{L}\|_\psi \le s$ for any $\mathcal{B} := \mathcal{A} * \mathcal{Z}$ with $\|\mathcal{Z}\|_\psi \le s$.*

The following lemma indicates that the ψ-RTEC(s) is equivalent to the ψ-USP(s) of $\mathcal{A}$.

Lemma 3.7. *The tensor $\mathcal{A}$ satisfies ψ-RTEC(s) if and only if it satisfies ψ-USP(s).*

Proof. $\Rightarrow$: Suppose that $\mathcal{A}$ satisfies ψ-RTEC(s). Assume on contradiction that the ψ-USP(s) is not satisfied, that is, there exists $\mathcal{B}$ such that there are two distinct tensors $\hat{\mathcal{L}}$ and $\tilde{\mathcal{L}}$ such that $\mathcal{A} * \hat{\mathcal{L}} = \mathcal{A} * \tilde{\mathcal{L}} = \mathcal{B}$ and $\|\hat{\mathcal{L}}\|_\psi = \|\tilde{\mathcal{L}}\|_\psi \le s$. Then $\mathcal{L} := \hat{\mathcal{L}} - \tilde{\mathcal{L}}$ is a tensor such that $\mathcal{A} * \mathcal{L} = 0$ and $\|\mathcal{L}\|_\psi \le 2s$, and thus $\zeta_{\min}(2s) = 0$, which is in contradiction with (3.7). Thus, $\mathcal{A}$ satisfies ψ-USP(s).

$\Leftarrow$: Suppose that $\mathcal{A}$ satisfies ψ-USP(s). Assume in contradiction that the ψ-RTEC(s) is not satisfied, that is, $\zeta_{\min}(2s) = 0$. Then by (3.7), there exists $\mathcal{L}$ with $\|\mathcal{L}\|_\psi \le 2s$ such that $\mathcal{A} * \mathcal{L} = 0$. Let $\mathcal{B} := \mathcal{A} * \mathcal{L}$, then it is easy to check that $\mathcal{A} * (\mathcal{L}/N) = \mathcal{A} * (\mathcal{L}/2N) = 0$ and $\|\mathcal{L}/2N\|_\psi \le \|\mathcal{L}/N\|_\psi \le s$ when N is sufficiently large. This contradicts the ψ-USP(s). Thus, $\mathcal{A}$ satisfies ψ-RTEC(s). ∎

Remark 3.8. If the ψ-RTEC(s) is not satisfied, one has no hope of recovering the ground truth s ψ-USP solution from the linear system $\mathcal{A} * \mathcal{L} = \mathcal{B}$.

In the following, we present some properties of the nonconvex function ψ, which are essential for the error bound analysis.

Theorem 3.9. *Suppose that $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are two arbitrary tensors. Then, the following properties hold:*

1. $\|\mathcal{B} - \mathcal{S}\|_\psi \ge \|\mathcal{B}\|_\psi - \|\mathcal{S}\|_\psi$;
2. $\|\mathcal{B} - \mathcal{S}\|_{\ell^\psi_{F,1}} \le \|\mathcal{B}\|_{\ell^\psi_{F,1}} + \|\mathcal{S}\|_{\ell^\psi_{F,1}}$;
3. $\psi(\|\mathcal{B}\|_F) \le \|\mathcal{B}\|_{\ell^\psi_{F,1}} \le \|\mathcal{B}\|_{\psi,1}$, *where* $\|\mathcal{B}\|_{\psi,1} := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \psi(|\mathcal{B}_{ijk}|)$.

The proof of Theorem 3.9 is given in Appendix C. Next, we develop an error bound for minimization problem (3.5).

Theorem 3.10. *Let $(\mathcal{L}^\natural, \mathcal{S}^\natural)$ be the pair of true low rank and sparse tensors, and $(\mathcal{A}^\star, \mathcal{L}^\star, \mathcal{S}^\star)$ be an optimal solution to the optimization problem (3.5). Assume that $\mathcal{A}^\star$ satisfies ψ-RTEC(s), $\mathcal{X} = \mathcal{A}^\star * \mathcal{L}^\natural + \mathcal{S}^\natural$, $\|\mathcal{L}^\natural\|_\psi \le \|\mathcal{L}^\star\|_\psi := s$, and $\lambda_2 > \lambda_1 r \vartheta^\psi_{r,s}$ with $r = \min\{n_1, n_2\}$. Then we have*

$$\psi\left(\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right) \le \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \le \frac{2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}}{\lambda_2 - \lambda_1 r \vartheta^\psi_{r,s}}, \tag{3.8}$$

where $\vartheta^\psi_{r,s}$ is a constant that depends on r, s, ψ.

Proof. By optimality, we have

$$\begin{aligned} &\sum_{u=1}^{3} \alpha_u \left\|\nabla_u \mathcal{A}^\star\right\|_\psi + \lambda_1 \left\|\mathcal{L}^\star\right\|_\psi + \lambda_2 \left\|\mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \\ &\quad \le \sum_{u=1}^{3} \alpha_u \left\|\nabla_u \mathcal{A}^\star\right\|_\psi + \lambda_1 \left\|\mathcal{L}^\natural\right\|_\psi + \lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}. \end{aligned} \tag{3.9}$$

Thus,

$$\lambda_2 \left\|\mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \le \lambda_1 \left(\left\|\mathcal{L}^\natural\right\|_\psi - \left\|\mathcal{L}^\star\right\|_\psi\right) + \lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}} \le \lambda_1 \left\|\mathcal{L}^\natural - \mathcal{L}^\star\right\|_\psi + \lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}, \tag{3.10}$$

where the last inequality follows from Theorem 3.9.1. By Theorem 3.9.2, we have

$$\lambda_2 \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \le \lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}} + \lambda_2 \left\|\mathcal{S}^\star\right\|_{\ell^\psi_{F,1}}. \tag{3.11}$$

Combining (3.10) and (3.11), we have

$$\lambda_2 \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \le \lambda_1 \left\|\mathcal{L}^\natural - \mathcal{L}^\star\right\|_\psi + 2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}. \tag{3.12}$$

Let $\mathcal{L}^\natural - \mathcal{L}^\star := \mathcal{U} * \mathcal{D} * \mathcal{V}^H$ and $\sigma^l_k := \sigma_l(\bar{D}^{(k)})$. By Assumption 3.1, one has

$$\begin{aligned} \left\|\mathcal{L}^\natural - \mathcal{L}^\star\right\|_\psi &= \sum_{k=1}^{n_3} \sum_{l=1}^{r} \frac{1}{n_3} \psi\left(\sigma^l_k\right) \le r\psi\left(\frac{1}{n_3 r} \sum_{k=1}^{n_3} \sum_{l=1}^{r} \sigma^l_k\right) \le r\psi\left(\sqrt{\frac{1}{n_3 r} \sum_{k=1}^{n_3} \sum_{l=1}^{r} \left(\sigma^l_k\right)^2}\right) \\ &= r\psi\left(r^{-\frac{1}{2}} \left\|\mathcal{D}\right\|_F\right) = r\psi\left(r^{-\frac{1}{2}} \left\|\mathcal{L}^\natural - \mathcal{L}^\star\right\|_F\right) \le r\psi\left(\frac{1}{\sqrt{r}\zeta_{\min}(2s)} \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right) \\ &\le r\vartheta^\psi_{r,s} \psi\left(\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right), \end{aligned} \tag{3.13}$$

where the first inequality follows from the concavity of ψ and Jensen's inequality, and the second inequality follows from the nondecreasing property of ψ and the mean inequality chain.

The third inequality uses $\|\mathcal{S}^\natural - \mathcal{S}^\star\|_F = \|\mathcal{A}^\star * (\mathcal{L}^\natural - \mathcal{L}^\star)\|_F \geq \zeta_{\min}(2s)\|\mathcal{L}^\natural - \mathcal{L}^\star\|_F$, $\vartheta^\psi_{r,s}$ is a constant related to r, s, ψ. Summing up (3.12) and (3.13), we notice that

$$\begin{aligned}\lambda_2 \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} &\leq \lambda_1 r \vartheta^\psi_{r,s} \psi\left(\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right) + 2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}} \\ &\leq \lambda_1 r \vartheta^\psi_{r,s} \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} + 2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}},\end{aligned}$$

where the last inequality uses Theorem 3.9.3. Provided that $\lambda_2 > \lambda_1 r \vartheta^\psi_{r,s}$, we get

$$\psi\left(\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right) \leq \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \leq \frac{2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}}{\lambda_2 - \lambda_1 r \vartheta^\psi_{r,s}}. \tag{3.14}$$

■

Remark 3.11. When the function ψ is given, $\vartheta^\psi_{r,s}$ is fixed. For example, when the function ψ takes the values of L1 and Lp, $\vartheta^\psi_{r,s}$ is equal to $\frac{1}{\sqrt{r}\zeta_{\min}(2s)}$ and $\frac{1}{\sqrt{r^p}\zeta^p_{\min}(2s)}$, respectively. From the definition of $\zeta_{\min}(2s)$ in (3.7), we can see that a smaller value of s, which corresponds to a lower rank of $\mathcal{L}^\star$, leads to a larger value of $\zeta_{\min}(2s)$. This means a smaller value of $\vartheta^\psi_{r,s}$, as $\zeta_{\min}(2s)$ and $\vartheta^\psi_{r,s}$ are inversely proportional. By Theorem 3.10, this demonstrates a smaller value of $\|\mathcal{S}^\natural - \mathcal{S}^\star\|_F$. The above analysis implies that the lower the rank of the background, the better the recovery effect.

Corollary 3.12. *When $\psi(x) = x^p$, the average of the entries of the sparse component $\mathcal{S}^\natural$ is bounded by T and the cardinality of the support $\mathcal{S}^\natural$ is bounded by m. By properly choosing λ_1 and λ_2, we have*

$$\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F \leq \sqrt[p]{4m}T.$$

Proof. When $\psi(x) = x^p$, one has

$$\left(\frac{\left\|\mathcal{S}^\natural\right\|^p_{\ell^\psi_{F,1}}}{m}\right)^{\frac{1}{p}} \leq \left(\frac{\left\|\mathcal{S}^\natural\right\|^p_{\psi,1}}{m}\right)^{\frac{1}{p}} \leq \frac{\left\|\mathcal{S}^\natural\right\|_1}{m} \leq T, \tag{3.15}$$

where the first inequality follows from Theorem 3.9.3, and the second inequality is because of the generalized power-mean inequality [5, Theorem 3]. Thus, (3.14) becomes

$$\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F \leq \left(\frac{2\lambda_2 m T^p}{\lambda_2 - \lambda_1 r \vartheta^\psi_{r,s}}\right)^{1/p}. \tag{3.16}$$

Take $\lambda_2 = 2\lambda_1 r \vartheta^\psi_{r,s}$, we have $\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F \leq \sqrt[p]{4m}T$. ■

Remark 3.13. Here we obtain the error bound similar to those in [61]. That is, for very sparse anomaly tensors, as long as T is bounded, then $\|\mathcal{S}^\natural - \mathcal{S}^\star\|_F/M$ with $M = n_1 n_2 n_3$ is rather small, indicating good recovery.

Although the error bound we obtain is very similar to that of [61], we show that in certain cases we get smaller error bounds.

Remark 3.14. In [61], if the transformation Γ satisfies $\mathcal{X}\circ\Gamma=\mathcal{X}$, then they derive an error bound based on Tucker rank as follows:

$$\left\|\mathcal{S}^{\natural}-\mathcal{S}^{\star}\right\|_F \leq \left(\frac{2\lambda_2 mT^p}{\lambda_2-\lambda_1\sum_{u=1}^{3}\alpha_u n_u^{1-\frac{p}{2}}}\right)^{1/p} := \text{error}_{tucker}. \tag{3.17}$$

In our model, if the dictionary $\mathcal{A}$ satisfies $\mathcal{A}=\mathcal{I}$, from Theorem 3.10, we obtain the error bound with $\psi(x)=x^p$ as follows:

$$\left\|\mathcal{S}^{\natural}-\mathcal{S}^{\star}\right\|_F \leq \left\|\mathcal{S}^{\natural}-\mathcal{S}^{\star}\right\|_{\ell^{\psi}_{F,1}}^{1/p} \leq \left(\frac{2\lambda_2 mT^p}{\lambda_2-\lambda_1 r^{1-\frac{p}{2}}}\right)^{1/p} := \text{error}_{tubal}. \tag{3.18}$$

Considering the majority of data sets used in HSIs, where the size satisfies the condition $n_3>\min\{n_1,n_2\}$, it can be derived that $r^{1-\frac{p}{2}}<\sum_{u=1}^{3}\alpha_u n_u^{1-\frac{p}{2}}$. This implies that our error bound is smaller than that in [61], i.e., $\text{error}_{tubal}<\text{error}_{tucker}$.

4. ELADMM algorithm and convergence analysis.

In this section, we propose an ELADMM algorithm to solve GNBRL model. Then the convergence of GNBRL is established. Finally, a stopping criterion for the algorithm is given.

4.1. ELADMM algorithm.

In real-world HSIs data, the entries in a spectral vector are corrupted by Gaussian noise [45, 62], so we convert (3.5) to the following problem,

$$\min_{\mathcal{A},\mathcal{L},\mathcal{S}} \sum_{u=1}^{3}\alpha_u\left\|\nabla_u\mathcal{A}\right\|_{\psi}+\lambda_1\left\|\mathcal{L}\right\|_{\psi}+\lambda_2\left\|\mathcal{S}\right\|_{\ell^{\psi}_{F,1}}+\beta f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right), \tag{4.1}$$

where $f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)=\frac{1}{2}\left\|\mathcal{A}*\mathcal{L}+\mathcal{S}-\mathcal{X}\right\|_F^2$. Now we develop an ELADMM algorithm to solve problem (4.1). By introducing the auxiliary variable $\mathcal{C}_u=\nabla_u\mathcal{A}$, $u\in[3]$, problem (4.1) can be rewritten as:

$$\min_{\mathcal{A},\mathcal{L},\mathcal{S},\{\mathcal{C}_u\}_{u=1}^{3}} \sum_{u=1}^{3}\alpha_u\left\|\mathcal{C}_u\right\|_{\psi}+\lambda_1\left\|\mathcal{L}\right\|_{\psi}+\lambda_2\left\|\mathcal{S}\right\|_{\ell^{\psi}_{F,1}}+\beta f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right) \quad \text{s.t.} \quad \mathcal{C}_u=\nabla_u\mathcal{A},\ u\in[3]. \tag{4.2}$$

The augmented Lagrangian function of (4.2) can be given by

$$\begin{aligned} L\left(\mathcal{S},\mathcal{A},\mathcal{C}_u,\mathcal{L};\mathcal{T}_u,\beta_u\right)=&\sum_{u=1}^{3}\left(\alpha_u\left\|\mathcal{C}_u\right\|_{\psi}+\left\langle\mathcal{T}_u,\nabla_u\mathcal{A}-\mathcal{C}_u\right\rangle+\frac{\beta_u}{2}\left\|\nabla_u\mathcal{A}-\mathcal{C}_u\right\|_F^2\right)\\ &+\lambda_1\left\|\mathcal{L}\right\|_{\psi}+\lambda_2\left\|\mathcal{S}\right\|_{\ell^{\psi}_{F,1}}+\beta f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right), \end{aligned} \tag{4.3}$$

where β_u for $u\in[3]$ are the penalty parameters, and $\mathcal{T}_u$ for $u\in[3]$ are the Lagrange multipliers. Under the ADMM algorithm framework, we can alternatively update each variable,

$$\begin{cases} \mathcal{S}^{t+1}\in\text{argmin}_{\mathcal{S}}\ L\left(\mathcal{S},\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t\right),\\ \mathcal{A}^{t+1}\in\text{argmin}_{\mathcal{A}}\ L\left(\mathcal{S}^{t+1},\mathcal{A},\mathcal{C}_u^t,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t\right),\\ \mathcal{C}_u^{t+1}\in\text{argmin}_{\mathcal{C}_u}\ L\left(\mathcal{S}^{t+1},\mathcal{A}^{t+1},\mathcal{C}_u,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t\right),\\ \mathcal{L}^{t+1}\in\text{argmin}_{\mathcal{L}}\ L\left(\mathcal{S}^{t+1},\mathcal{A}^{t+1},\mathcal{C}_u^{t+1},\mathcal{L};\mathcal{T}_u^t,\beta_u^t\right),\\ \mathcal{T}_u^{t+1}=\mathcal{T}_u^t+\beta_u^t\left(\nabla_u\mathcal{A}^{t+1}-\mathcal{C}_u^{t+1}\right),\ \beta_u^{t+1}=\rho\beta_u^t, \end{cases}$$

where t denotes the iteration number and ρ is a constant value greater than 1. We then shall discuss how to solve its subproblems for each involved variable.

4.1.1. Computing $\mathcal{S}^{t+1}$. Fixing other variables except for $\mathcal{S}$ in (4.3), $\mathcal{S}^{t+1}$ is determined by solving the following optimization problem:

$$\arg\min_{\mathcal{S}} \ \lambda_2 \|\mathcal{S}\|_{\ell^{\psi}_{F,1}} + \frac{\beta}{2}\left\|\mathcal{A}^t * \mathcal{L}^t + \mathcal{S} - \mathcal{X}\right\|_F^2. \tag{4.4}$$

To solve this problem, we need to introduce the following theorem.

Theorem 4.1. *For a positive number λ, the proximal operator of $\operatorname{prox}_{\lambda\|\cdot\|_F^{\psi}}(\boldsymbol{z})$ has a closed-form solution, i.e.,*

$$\boldsymbol{x}^\star = \operatorname{prox}_{\lambda\|\cdot\|_F^{\psi}}(\boldsymbol{z}) := \arg\min_{\boldsymbol{x}} \left\{\lambda\psi\left(\|\boldsymbol{x}\|_F\right) + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_F^2\right\} = \begin{cases} \phi\left(\|\boldsymbol{z}\|_F\right)\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_F}, & \boldsymbol{z}\neq 0, \\ 0, & \boldsymbol{z}=0, \end{cases}$$

where

$$\phi(z) \in \operatorname{prox}_{\lambda\psi}(z) := \arg\min_{x\in\mathbb{R}_+}\left\{\lambda\psi(x) + \frac{1}{2}(x-z)^2\right\}.$$

Proof. It is clear that $\boldsymbol{x}^\star = 0$ when $\boldsymbol{z} = 0$. Therefore, we only need to consider $\boldsymbol{z}\neq 0$ in the following. From $\phi(z)\in\operatorname{prox}_{\lambda\psi}(z)$, we have

$$\begin{aligned}
&\lambda\psi\left(\left\|\phi\left(\|\boldsymbol{z}\|_F\right)\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_F}\right\|_F\right) + \frac{1}{2}\left\|\phi(\|\boldsymbol{z}\|_F)\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_F} - \boldsymbol{z}\right\|_F^2 = \lambda\psi\left(\phi\left(\|\boldsymbol{z}\|_F\right)\right) + \frac{1}{2}\left(\phi\left(\|\boldsymbol{z}\|_F\right) - \|\boldsymbol{z}\|_F\right)^2 \\
&\leq \lambda\psi\left(\|\boldsymbol{x}\|_F\right) + \frac{1}{2}\left(\|\boldsymbol{x}\|_F - \|\boldsymbol{z}\|_F\right)^2 \leq \lambda\psi\left(\|\boldsymbol{x}\|_F\right) + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_F^2,
\end{aligned}$$

which completes the proof. ∎

By Theorem 4.1, the optimal solution of (4.4) is given by

$$\mathcal{S}^{t+1}(i,j,:) = \operatorname{prox}_{\frac{\lambda_2}{\beta}\psi}\left(\|\boldsymbol{z_{ij}}\|_F\right)\frac{\boldsymbol{z_{ij}}}{\|\boldsymbol{z_{ij}}\|_F}, \tag{4.5}$$

where $\boldsymbol{z_{ij}} = \mathcal{Z}(i,j,:)$ with $\mathcal{Z} = \mathcal{X} - \mathcal{A}^t * \mathcal{L}^t$ for $i\in[n_1]$, $j\in[n_2]$.

4.1.2. Computing $\mathcal{A}^{t+1}$. Although $L(\mathcal{S},\mathcal{A},\mathcal{C}_u,\mathcal{L};\mathcal{T}_u,\beta_u)$ with respect to $\mathcal{A}$ is convex, there is no closed-form solution. Similarly to [51, 57], we update $\mathcal{A}$ by solving the following subproblem,

$$\begin{aligned}
&\arg\min_{\mathcal{A}} \sum_{u=1}^{3}\left(\left\langle\mathcal{T}_u^t, \nabla_u\mathcal{A} - \mathcal{C}_u^t\right\rangle + \frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A} - \mathcal{C}_u^t\right\|_F^2\right) + \beta\left\langle\nabla_{\mathcal{A}} f\left(\hat{\mathcal{A}}^t, \mathcal{L}^t, \mathcal{S}^{t+1}\right), \mathcal{A} - \hat{\mathcal{A}}^t\right\rangle \\
&\quad + \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A} - \hat{\mathcal{A}}^t\right\|_F^2 \\
&= \arg\min_{\mathcal{A}} \sum_{u=1}^{3}\frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A} - \left(\mathcal{C}_u^t - \frac{\mathcal{T}_u^t}{\beta_u^t}\right)\right\|_F^2 + \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A} - \left(\hat{\mathcal{A}}^t - \frac{\nabla_{\mathcal{A}} f\left(\hat{\mathcal{A}}^t, \mathcal{L}^t, \mathcal{S}^{t+1}\right)}{l_{\mathcal{A}}^t}\right)\right\|_F^2,
\end{aligned} \tag{4.6}$$

where $l_{\mathcal{A}}^t \geq l_{\mathcal{A}}(f)$, $l_{\mathcal{A}}(f)$ is a Lipschitz constant of $\nabla_{\mathcal{A}} f(\mathcal{A}, \mathcal{L}^t, \mathcal{S}^{t+1}) = (\mathcal{A} * \mathcal{L}^t + \mathcal{S}^{t+1} - \mathcal{X}) * (\mathcal{L}^t)^H$ with respect to $\mathcal{A}$, and $\hat{\mathcal{A}}^t$ is an extrapolated point. To optimize the above problem, we can solve the following linear system,

$$\left(\beta l_{\mathcal{A}}^t \mathcal{I} + \sum_{u=1}^{3} \beta_u^t \nabla_u^H \nabla_u\right)(\mathcal{A}) = \beta l_{\mathcal{A}}^t \hat{\mathcal{A}}^t - \beta \nabla_{\mathcal{A}} f\left(\hat{\mathcal{A}}^t, \mathcal{L}^t, \mathcal{S}^{t+1}\right) + \sum_{u=1}^{3} \nabla_u^H \left(\beta_u^t \mathcal{C}_u^t - \mathcal{T}_u^t\right), \tag{4.7}$$

where $\nabla_u^H(\cdot)$ denotes the transpose operator of $\nabla_u(\cdot)$. Due to the block-circulant structure of the matrix corresponding to the operator $\nabla_u^H \nabla_u$, it can be diagonalized by the three-imensional FFT matrix. Specifically, following [26, 39], by performing a Fourier transform on both sides of (4.7) and adopting the convolution theorem, the closed-form solution to $\mathcal{A}^{t+1}$ can be easily deduced as

$$\mathcal{A}^{t+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\beta l_{\mathcal{A}}^t \hat{\mathcal{A}}^t - \beta \nabla_{\mathcal{A}} f\left(\hat{\mathcal{A}}^t, \mathcal{L}^t, \mathcal{S}^{t+1}\right)\right) + \sum_{u=1}^{3} \mathcal{F}(\nabla_u)^H \odot \mathcal{F}\left(\beta_u^t \mathcal{C}_u^t - \mathcal{T}_u^t\right)}{\beta l_{\mathcal{A}}^t \mathcal{E} + \sum_{u=1}^{3} \beta_u^t \left|\mathcal{F}(\nabla_u)\right|^2}\right). \tag{4.8}$$

Here $\mathcal{E}$ represents the tensor with all elements as 1, $\odot$ is the elementwise multiplication, $\mathcal{F}(\cdot)$ is the Fourier transform, and $|\cdot|^2$ is the elementwise square operation. Here, we take $l_{\mathcal{A}}^t = \|\mathcal{L}^t\|^2 + \epsilon$ with $\epsilon > 0$ and $\hat{\mathcal{A}}^t = \mathcal{A}^t + \omega_{\mathcal{A}}^t(\mathcal{A}^t - \mathcal{A}^{t-1})$, where

$$\omega_{\mathcal{A}}^t = \min\left\{\hat{\omega}^t, \delta_\omega \sqrt{\frac{l_{\mathcal{A}}^{t-1}}{l_{\mathcal{A}}^t}}\right\}, \delta_\omega < 1, \hat{\omega}^t = \frac{l^{t-1} - 1}{l^t}, l^0 = 1, l^t = \frac{1}{2}\left(1 + \sqrt{1 + 4\left(l^{t-1}\right)^2}\right).$$

4.1.3. Computing $\mathcal{C}_u^{t+1}$. Fixing other variables except for $\mathcal{C}_u$ in (4.3), we update $\mathcal{C}_u$ by solving the following subproblem:

$$\begin{aligned} &\arg\min_{\mathcal{C}_u} \ \alpha_u \|\mathcal{C}_u\|_\psi + \left\langle \mathcal{T}_u^t, \nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u \right\rangle + \frac{\beta_u^t}{2} \left\|\nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u\right\|_F^2 \\ &= \arg\min_{\mathcal{C}_u} \ \alpha_u \|\mathcal{C}_u\|_\psi + \frac{\beta_u^t}{2} \left\|\mathcal{C}_u - \left(\nabla_u \mathcal{A}^{t+1} + \frac{\mathcal{T}_u^t}{\beta_u^t}\right)\right\|_F^2. \end{aligned} \tag{4.9}$$

By Lemma B.1, (4.9) can be solved by

$$\mathcal{C}_u^{t+1} = \operatorname{prox}_{\frac{\alpha_u}{\beta_u^t}\|\cdot\|_\psi}\left(\nabla_u \mathcal{A}^{t+1} + \frac{\mathcal{T}_u^t}{\beta_u^t}\right). \tag{4.10}$$

4.1.4. Computing $\mathcal{L}^{t+1}$. Fixing other variables except for $\mathcal{L}$ in (4.3), we update $\mathcal{L}$ by solving the following subproblem:

$$\arg\min_{\mathcal{L}} \ \lambda_1 \|\mathcal{L}\|_\psi + \beta f\left(\mathcal{A}^{t+1}, \mathcal{L}, \mathcal{S}^{t+1}\right). \tag{4.11}$$

Since the $\|\mathcal{L}\|_\psi$ regularizer is nonconvex, we cannot use an extrapolation method like the one for computing $\mathcal{A}$. We propose a new extrapolation method to solve problem (4.11). First, we

give an extrapolated point $\hat{\mathcal{L}}^t = \mathcal{L}^t + \omega_{\mathcal{L}}^t(\mathcal{L}^t - \mathcal{L}^{t-1})$ with $\omega_{\mathcal{L}}^t = \min\{\hat{\omega}^t, \frac{\delta_\omega(\gamma-1)}{2\gamma}\sqrt{l_{\mathcal{L}}^{t-1}/l_{\mathcal{L}}^t}\}$ and $l_{\mathcal{L}}^t = \gamma\|\mathcal{A}^{t+1}\|^2 + \epsilon$, $\gamma > 1$. Then we obtain $\mathcal{L}^{t+1}$ by solving the following problem:

$$\begin{aligned} &\arg\min_{\mathcal{L}} \ \lambda_1 \|\mathcal{L}\|_\psi + \beta\left\langle \nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \hat{\mathcal{L}}^t, \mathcal{S}^{t+1}\right), \mathcal{L} - \hat{\mathcal{L}}^t\right\rangle + \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L} - \hat{\mathcal{L}}^t\right\|_F^2 \\ &= \arg\min_{\mathcal{L}} \ \lambda_1 \|\mathcal{L}\|_\psi + \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L} - \left(\hat{\mathcal{L}}^t - \frac{\nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \hat{\mathcal{L}}^t, \mathcal{S}^{t+1}\right)}{l_{\mathcal{L}}^t}\right)\right\|_F^2. \end{aligned} \tag{4.12}$$

By Lemma B.1, the optimal solution of (4.12) is

$$\mathcal{L}^{t+1} = \operatorname{prox}_{\frac{\lambda_1}{\beta l_{\mathcal{L}}^t}\|\cdot\|_\psi}\left(\hat{\mathcal{L}}^t - \frac{\nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \hat{\mathcal{L}}^t, \mathcal{S}^{t+1}\right)}{l_{\mathcal{L}}^t}\right). \tag{4.13}$$

We summarize the algorithm for GNBRL in Algorithm 4.1.

Remark 4.2. We choose ADMM for optimization based on the reformulation in (4.2). To enable efficient computation of $\mathcal{A}^{t+1}$ and $\mathcal{L}^{t+1}$, we apply a linearization technique to the smooth part $f(\mathcal{A}, \mathcal{L}, \mathcal{S})$ with an additional proximal term to split the interdependent term such that they can be solved independently. Moreover, we incorporate an extrapolation method in Steps 2 and 5 to improve the performance of the algorithm. In summary, we develop an ELADMM algorithm to solve our proposed GNBRL model.

Algorithm 4.1 ELADMM method to solve (4.2).

Input: The tensor data $\mathcal{X}$, parameters $\{\alpha_u\}_{u=1}^3$, λ_1, λ_2, β.
Initialize: $\mathcal{A}^0$, $\mathcal{L}^0$, $\mathcal{S}^0$, $\{\mathcal{C}_u^0, \mathcal{T}_u^0\}_{u=1}^3$, $\{\beta_u^0\}_{u=1}^3$.
while not converge **do**
***Step* 1**. Update $\mathcal{S}^{t+1}$ according to (4.5).
***Step* 2**. Let $\hat{\mathcal{A}}^t = \mathcal{A}^t + \omega_{\mathcal{A}}^t(\mathcal{A}^t - \mathcal{A}^{t-1})$.
***Step* 3**. Update $\mathcal{A}^{t+1}$ according to (4.8).
***Step* 4**. Update $\mathcal{C}_u^{t+1}$ according to (4.10).
***Step* 5**. Let $\hat{\mathcal{L}}^t = \mathcal{L}^t + \omega_{\mathcal{L}}^t(\mathcal{L}^t - \mathcal{L}^{t-1})$.
***Step* 6**. Update $\mathcal{L}^{t+1}$ according to (4.13).
***Step* 7**. Update multipliers $\mathcal{T}_u^{t+1}$ and penalty parameters β_u^{t+1} according to

$$\begin{cases} \mathcal{T}_u^{t+1} = \mathcal{T}_u^t + \beta_u^t\left(\nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right), \\ \beta_u^{t+1} = \rho\beta_u^t, \ u \in [3]. \end{cases} \tag{4.14}$$

Let $t := t + 1$ and go to ***Step* 1**.
end while
Output: $\mathcal{S}^{t+1}$, $\mathcal{A}^{t+1}$, $\mathcal{L}^{t+1}$.

4.2. Convergence analysis. This subsection presents the convergence analysis of our proposed algorithm. We state the main results in Theorem 4.7 below. Before that, we present some lemmas.

Lemma 4.3. *The sequence* $\{\mathcal{T}_u^t\}_{t=1}^{\infty}$ *for each* $u \in [3]$ *is bounded.*

Proof. By the first order necessary optimality condition of (4.9), we have

$$0 \in \alpha_u \frac{\partial}{\partial \mathcal{C}_u}\Big|_{\mathcal{C}_u=\mathcal{C}_u^{t+1}} \|\mathcal{C}_u\|_\psi - \mathcal{T}_u^t + \beta_u^t \left(C_u^{t+1} - \nabla_u \mathcal{A}^{t+1}\right). \tag{4.15}$$

Let $\mathcal{C}_u = \mathcal{U} * \mathcal{D} * \mathcal{V}^H$. Then it follows that

$$\frac{\partial}{\partial \bar{C}_u^{(k)}} \left\|\bar{C}_u^{(k)}\right\|_\psi = \bar{U}^{(k)} \operatorname{diag}\left(\partial\psi\left(\bar{D}_{i,i}^{(k)}\right)\right)\left(\bar{V}^{(k)}\right)^H,$$

and then one can obtain

$$\left\| \frac{\partial}{\partial \bar{C}_u^{(k)}} \left\|\bar{C}_u^{(k)}\right\|_\psi \right\|_F = \sum_{i=1}^{r} \partial\psi\left(\bar{D}_{i,i}^{(k)}\right).$$

Next, we will prove that $\sum_{i=1}^r \partial\psi(\bar{D}_{i,i}^{(k)})$ is bounded. We split the proof into five cases.

Case 1. $\psi(x) = x$. $\sum_{i=1}^r \partial\psi(\bar{D}_{i,i}^{(k)}) = r$ is bounded.

Case 2. $\psi(x) = x^p$. In order to overcome the singularity of $(x^p)' = p/x^{1-p}$ near ∞ as x near 0, we consider for $0 < \epsilon \ll 1$ the approximation

$$(x^p)' \approx \frac{p}{\max\{\epsilon^{1-p}, x^{1-p}\}}.$$

Thus,

$$\sum_{i=1}^{r} \partial\psi\left(\bar{D}_{i,i}^{(k)}\right) = \sum_{i=1}^{r} \frac{p}{\max\left\{\epsilon^{1-p}, \left(\bar{D}_{i,i}^{(k)}\right)^{1-p}\right\}} \le \frac{pr}{\epsilon^{1-p}}$$

is bounded.

Case 3. $\psi(x) = \psi^{\mathrm{MCP}}(x)$. $\sum_{i=1}^r \partial\psi(\bar{D}_{i,i}^{(k)}) \le r$ is bounded.

Case 4. $\psi(x) = \log(\frac{x}{\theta} + 1)$.

$$\sum_{i=1}^{r} \partial\psi\left(\bar{D}_{i,i}^{(k)}\right) = \sum_{i=1}^{r} \frac{1}{\bar{D}_{i,i}^{(k)} + \theta} \le \frac{r}{\theta}$$

is bounded.

Case 5. $\psi \in \{\psi^{\mathrm{CapL1}}, \psi^{\mathrm{CapLp}}, \psi^{\mathrm{CapMCP}}, \psi^{\mathrm{CapLog}}\}$. From Cases 1–4, we obtain the boundedness of $\sum_{i=1}^r \partial\psi\left(\bar{D}_{i,i}^{(k)}\right)$.

Combining all cases, we see that

$$\frac{\partial}{\partial \bar{\mathcal{C}}_u} \|\mathcal{C}_u\|_\psi = \left[\frac{\partial}{\partial \bar{C}_u^{(1)}} \left\|\bar{C}_u^{(1)}\right\|_\psi | \cdots | \frac{\partial}{\partial \bar{C}_u^{(n_3)}} \left\|\bar{C}_u^{(n_3)}\right\|_\psi\right]$$

is bounded. For $\bar{\mathcal{C}}_u = \mathcal{C}_u \times_3 F_{n_3}$ and using the chain rule in matrix calculus, one can obtain that $\frac{\partial}{\partial \mathcal{C}_u}\|\mathcal{C}_u\|_\psi = \frac{\partial}{\partial \bar{\mathcal{C}}_u}\|\mathcal{C}_u\|_\psi \times_3 F_{n_3}^H$ is bounded. Combining this with (4.14) and (4.15),

$$\mathcal{T}_u^{t+1} \in \alpha_u \frac{\partial}{\partial \mathcal{C}_u}\bigg|_{\mathcal{C}_u=\mathcal{C}_u^{t+1}} \|\mathcal{C}_u\|_\psi$$

is bounded. ∎

Lemma 4.4. *Let $\{\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\}_{t=1}^\infty$ be the sequence generated by Algorithm 4.1, then*

$$\begin{aligned} &L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}^{t+1}; \mathcal{T}_u^t, \beta_u^t\right) - L\left(\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right) \\ &\le \frac{\beta l_{\mathcal{A}}^{t-1}}{2}\delta_\omega^2 \left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}^{t+1} - \mathcal{A}^t\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\delta_\omega^2 \left\|\mathcal{L}^t - \mathcal{L}^{t-1}\right\|_F^2 \\ &\quad - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{4\gamma}\left\|\mathcal{L}^{t+1} - \mathcal{L}^t\right\|_F^2. \end{aligned}$$

Proof. Since $\mathcal{S}^{t+1}$ is optimal to (4.4), we have

$$L\left(\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right) \ge L\left(\mathcal{S}^{t+1}, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right). \tag{4.16}$$

Using Lemma 2.1 of [54], we have

$$\begin{aligned} &L\left(\mathcal{S}^{t+1}, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right) - L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right) \\ &\ge \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}^{t+1} - \hat{\mathcal{A}}^t\right\|_F^2 + \beta l_{\mathcal{A}}^t\left\langle \hat{\mathcal{A}}^t - \mathcal{A}^t, \mathcal{A}^{t+1} - \hat{\mathcal{A}}^t\right\rangle \\ &= \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}^{t+1} - \mathcal{A}^t\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^t}{2}\left(\omega_{\mathcal{A}}^t\right)^2\left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2 \\ &\ge \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}^{t+1} - \mathcal{A}^t\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^{t-1}}{2}\delta_\omega^2\left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2, \end{aligned} \tag{4.17}$$

where we have used $\omega_{\mathcal{A}}^t \le \delta_\omega\sqrt{l_{\mathcal{A}}^{t-1}/l_{\mathcal{A}}^t}$ to get the last inequality.

Since $\mathcal{C}_u^{t+1}$ is optimal to (4.10), we have

$$\begin{aligned} &\alpha_u\left\|\mathcal{C}_u^{t+1}\right\|_\psi + \left\langle \mathcal{T}_u^t, \nabla_u\mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right\rangle + \frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right\|_F^2 \\ &\le \alpha_u\left\|\mathcal{C}_u^t\right\|_\psi + \left\langle \mathcal{T}_u^t, \nabla_u\mathcal{A}^{t+1} - \mathcal{C}_u^t\right\rangle + \frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A}^{t+1} - \mathcal{C}_u^t\right\|_F^2, \end{aligned} \tag{4.18}$$

which implies

$$L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right) \le L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right). \tag{4.19}$$

Denote $\nabla_{\mathcal{L}} f_{\mathcal{L}} := \nabla_{\mathcal{L}} f(\mathcal{A}^{t+1}, \mathcal{L}, \mathcal{S}^{t+1})$. Since $\mathcal{L}^{t+1}$ is optimal to (4.12), we have

$$\begin{aligned} &\lambda_1\left\|\mathcal{L}^{t+1}\right\|_\psi + \beta\left\langle \nabla_{\mathcal{L}} f_{\hat{\mathcal{L}}^t}, \mathcal{L}^{t+1} - \hat{\mathcal{L}}^t\right\rangle + \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L}^{t+1} - \hat{\mathcal{L}}^t\right\|_F^2 \\ &\le \lambda_1\left\|\mathcal{L}^t\right\|_\psi + \beta\left\langle \nabla_{\mathcal{L}} f_{\hat{\mathcal{L}}^t}, \mathcal{L}^t - \hat{\mathcal{L}}^t\right\rangle + \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L}^t - \hat{\mathcal{L}}^t\right\|_F^2. \end{aligned} \tag{4.20}$$

Since $\nabla_{\mathcal{L}} f_{\mathcal{L}}$ is Lipschitz continuous, we have

$$f\left(\mathcal{A}^{t+1},\mathcal{L}^{t+1},\mathcal{S}^{t+1}\right) \le f\left(\mathcal{A}^{t+1},\mathcal{L}^{t},\mathcal{S}^{t+1}\right) + \left\langle \nabla_{\mathcal{L}} f_{\mathcal{L}^t}, \mathcal{L}^{t+1}-\mathcal{L}^t\right\rangle + \frac{l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2. \tag{4.21}$$

Summing up (4.20) and (4.21), we notice that

$$\begin{aligned}
&\lambda_1\left\|\mathcal{L}^{t+1}\right\|_{\psi} + \beta f\left(\mathcal{A}^{t+1},\mathcal{L}^{t+1},\mathcal{S}^{t+1}\right) - \lambda_1\left\|\mathcal{L}^{t}\right\|_{\psi} - \beta f\left(\mathcal{A}^{t+1},\mathcal{L}^{t},\mathcal{S}^{t+1}\right)\\
&\le \beta\left\langle \nabla_{\mathcal{L}} f_{\mathcal{L}^t} - \nabla_{\mathcal{L}} f_{\hat{\mathcal{L}}^t}, \mathcal{L}^{t+1}-\mathcal{L}^t\right\rangle + \frac{\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2 + \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L}^t-\hat{\mathcal{L}}^t\right\|_F^2 - \frac{\beta l_{\mathcal{L}}^t}{2}\left\|\mathcal{L}^{t+1}-\hat{\mathcal{L}}^t\right\|_F^2\\
&= \beta\left\langle \nabla_{\mathcal{L}} f_{\mathcal{L}^t} - \nabla_{\mathcal{L}} f_{\hat{\mathcal{L}}^t}, \mathcal{L}^{t+1}-\mathcal{L}^t\right\rangle + \beta l_{\mathcal{L}}^t\left\langle \mathcal{L}^t-\mathcal{L}^{t+1}, \mathcal{L}^t-\hat{\mathcal{L}}^t\right\rangle - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&= \beta\left\langle \mathcal{L}^{t+1}-\mathcal{L}^t, \nabla_{\mathcal{L}} f_{\mathcal{L}^t} - \nabla_{\mathcal{L}} f_{\hat{\mathcal{L}}^t} - l_{\mathcal{L}}^t\left(\mathcal{L}^t-\hat{\mathcal{L}}^t\right)\right\rangle - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&= \beta\left\langle \mathcal{L}^{t+1}-\mathcal{L}^t, \left(\left(\mathcal{A}^{t+1}\right)^H * \mathcal{A}^{t+1} - l_{\mathcal{L}}^t\mathcal{I}\right) * \left(\mathcal{L}^t-\hat{\mathcal{L}}^t\right)\right\rangle - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&\le \beta\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F\left\|\mathcal{L}^t-\hat{\mathcal{L}}^t\right\|_F\left\|\left(\mathcal{A}^{t+1}\right)^H * \mathcal{A}^{t+1} - l_{\mathcal{L}}^t\mathcal{I}\right\| - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&\le \beta l_{\mathcal{L}}^t\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F\left\|\mathcal{L}^t-\hat{\mathcal{L}}^t\right\|_F - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&= \beta l_{\mathcal{L}}^t\omega_{\mathcal{L}}^t\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{2\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&\le \frac{\gamma\beta l_{\mathcal{L}}^t}{(\gamma-1)}\left(\omega_{\mathcal{L}}^t\right)^2\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2 - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{4\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2\\
&\le \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\delta_\omega^2\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2 - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{4\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2.
\end{aligned} \tag{4.22}$$

Here, we have used the Cauchy–Schwarz inequality in the second inequality, the Young's inequality in the fourth one, and $\omega_{\mathcal{L}}^t \le \frac{\delta_\omega(\gamma-1)}{2\gamma}\sqrt{l_{\mathcal{L}}^{t-1}/l_{\mathcal{L}}^t}$ to get the last inequality, the fact $\hat{\mathcal{L}}^t = \mathcal{L}^t + \omega_{\mathcal{L}}^t(\mathcal{L}^t-\mathcal{L}^{t-1})$ to have the last equality. (4.22) implies that

$$\begin{aligned}
&L\left(\mathcal{S}^{t+1},\mathcal{A}^{t+1},\mathcal{C}_u^{t+1},\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t\right) - L\left(\mathcal{S}^{t+1},\mathcal{A}^{t+1},\mathcal{C}_u^{t+1},\mathcal{L}^{t+1};\mathcal{T}_u^t,\beta_u^t\right)\\
&\ge \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{4\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2 - \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\delta_\omega^2\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2.
\end{aligned} \tag{4.23}$$

Summing up (4.16), (4.17), (4.19), and (4.23), one has

$$\begin{aligned}
&L\left(\mathcal{S}^{t+1},\mathcal{A}^{t+1},\mathcal{C}_u^{t+1},\mathcal{L}^{t+1};\mathcal{T}_u^t,\beta_u^t\right) - L\left(\mathcal{S}^{t},\mathcal{A}^{t},\mathcal{C}_u^{t},\mathcal{L}^{t};\mathcal{T}_u^t,\beta_u^t\right)\\
&\le \frac{\beta l_{\mathcal{A}}^{t-1}}{2}\delta_\omega^2\left\|\mathcal{A}^t-\mathcal{A}^{t-1}\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}^{t+1}-\mathcal{A}^t\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\delta_\omega^2\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2\\
&\quad - \frac{(\gamma-1)\beta l_{\mathcal{L}}^t}{4\gamma}\left\|\mathcal{L}^{t+1}-\mathcal{L}^t\right\|_F^2.
\end{aligned}$$

Lemma 4.5. *The sequence* $\{\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t, \mathcal{T}_u^t\}_{t=1}^{\infty}$ *generated by Algorithm 4.1 is bounded if the sequence* $\{\mathcal{A}^t, \mathcal{L}^t\}_{t=1}^{\infty}$ *is bound.*

Proof. According to the iterative process of Algorithm 4.1 we can obtain

$$\begin{aligned} &L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}^{t+1}; \mathcal{T}_u^{t+1}, \beta_u^{t+1}\right) - L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}^{t+1}; \mathcal{T}_u^{t}, \beta_u^{t}\right) \\ &= \sum_{u=1}^{3}\left(\left\langle \mathcal{T}_u^{t+1} - \mathcal{T}_u^t, \nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right\rangle + \frac{\beta_u^{t+1} - \beta_u^t}{2}\left\|\nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right\|_F^2\right) \\ &= \sum_{u=1}^{3}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2}\left\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\right\|_F^2. \end{aligned} \tag{4.24}$$

Combining this with Lemma 4.4 gives

$$\begin{aligned} &L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}^{t+1}; \mathcal{T}_u^{t+1}, \beta_u^{t+1}\right) - L\left(\mathcal{S}^{t}, \mathcal{A}^{t}, \mathcal{C}_u^{t}, \mathcal{L}^{t}; \mathcal{T}_u^{t}, \beta_u^{t}\right) \\ &\le \frac{\beta l_{\mathcal{A}}^{t-1}}{2}\delta_\omega^2\left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^{t}}{2}\left\|\mathcal{A}^{t+1} - \mathcal{A}^{t}\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\delta_\omega^2\left\|\mathcal{L}^t - \mathcal{L}^{t-1}\right\|_F^2 \\ &\quad - \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t}}{4\gamma}\left\|\mathcal{L}^{t+1} - \mathcal{L}^{t}\right\|_F^2 + \sum_{u=1}^{3}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2}\left\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\right\|_F^2 \\ &= \frac{\beta l_{\mathcal{A}}^{t-1}}{2}\left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2 - \frac{\beta l_{\mathcal{A}}^{t}}{2}\left\|\mathcal{A}^{t+1} - \mathcal{A}^{t}\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\left\|\mathcal{L}^t - \mathcal{L}^{t-1}\right\|_F^2 \\ &\quad - \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t}}{4\gamma}\left\|\mathcal{L}^{t+1} - \mathcal{L}^{t}\right\|_F^2 + \frac{\beta l_{\mathcal{A}}^{t-1}\left(\delta_\omega^2 - 1\right)}{2}\left\|\mathcal{A}^t - \mathcal{A}^{t-1}\right\|_F^2 \\ &\quad + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{t-1}}{4\gamma}\left(\delta_\omega^2 - 1\right)\left\|\mathcal{L}^t - \mathcal{L}^{t-1}\right\|_F^2 + \sum_{u=1}^{3}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2}\left\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\right\|_F^2. \end{aligned} \tag{4.25}$$

Then, we have

$$\begin{aligned} &L\left(\mathcal{S}^{T+1}, \mathcal{A}^{T+1}, \mathcal{C}_u^{T+1}, \mathcal{L}^{T+1}; \mathcal{T}_u^{T+1}, \beta_u^{T+1}\right) - L\left(\mathcal{S}^{1}, \mathcal{A}^{1}, \mathcal{C}_u^{1}, \mathcal{L}^{1}; \mathcal{T}_u^{1}, \beta_u^{1}\right) \\ &\le \frac{\beta l_{\mathcal{A}}^{0}}{2}\left\|\mathcal{A}^1 - \mathcal{A}^0\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{0}}{4\gamma}\left\|\mathcal{L}^1 - \mathcal{L}^0\right\|_F^2 + \sum_{t=1}^{T}\sum_{u=1}^{3}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2}\left\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\right\|_F^2 \\ &\le \frac{\beta l_{\mathcal{A}}^{0}}{2}\left\|\mathcal{A}^1 - \mathcal{A}^0\right\|_F^2 + \frac{(\gamma-1)\beta l_{\mathcal{L}}^{0}}{4\gamma}\left\|\mathcal{L}^1 - \mathcal{L}^0\right\|_F^2 + \sum_{u=1}^{3}\max_t\left\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\right\|_F^2\sum_{t=1}^{T}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2}. \end{aligned}$$

Given that $\{\mathcal{T}_u^t\}_{t=1}^{\infty}$ is bounded, the quantity $\max_t\|\mathcal{T}_u^{t+1} - \mathcal{T}_u^t\|_F^2$ must be bounded as well. Notice that $\beta_u^{t+1} = \rho\beta_u^t = \rho^t\beta_u^1$, then

$$\sum_{t=1}^{\infty}\frac{\beta_u^t + \beta_u^{t+1}}{2\left(\beta_u^t\right)^2} = \frac{1+\rho}{2\beta_u^1}\sum_{t=1}^{\infty}\frac{1}{\rho^{t-1}} = \frac{\rho(\rho+1)}{2\beta_u^1(\rho-1)}$$

is bounded, and hence $L(\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t)$ is upper bounded. We have from (4.3) that

(4.26)
$$\begin{aligned}&L\left(\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t\right)+\sum_{u=1}^{3}\frac{1}{2\beta_u^t}\left\|\mathcal{T}_u^t\right\|_F^2\\&=\sum_{u=1}^{3}\left(\alpha_u\left\|\mathcal{C}_u^t\right\|_\psi+\frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A}^t-\mathcal{C}_u^t+\frac{\mathcal{T}_u^t}{\beta_u^t}\right\|_F^2\right)+\lambda_1\left\|\mathcal{L}^t\right\|_\psi+\lambda_2\left\|\mathcal{S}^t\right\|_{\ell_{F,1}^\psi}+\beta f\left(\mathcal{A}^t,\mathcal{L}^t,\mathcal{S}^t\right).\end{aligned}$$

Note that each term on the right-hand side of (4.26) is nonnegative. By the boundedness of $\mathcal{T}_u^t$ for $u\in[3]$ and $L(\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t;\mathcal{T}_u^t,\beta_u^t)$, we know that $\|\nabla_u\mathcal{A}^t-\mathcal{C}_u^t+\mathcal{T}_u^t/\beta_u^t\|_F$ and $\|\mathcal{A}^t*\mathcal{L}^t+\mathcal{S}^t-\mathcal{X}\|_F$ are bounded. Benefiting from bounded $\mathcal{A}^t$, $\mathcal{L}^t$, and $\mathcal{T}_u^t$, we can conclude that $\mathcal{S}^t$ and $\mathcal{C}_u^t$ have a boundary. ∎

Lemma 4.6. *The sequence $\{\mathcal{A}^t,\mathcal{L}^t\}_{t=1}^\infty$ generated by Algorithm 4.1 satisfies $\lim_{t\to\infty}(\mathcal{A}^t-\mathcal{A}^{t-1})=0$ and $\lim_{t\to\infty}(\mathcal{L}^t-\mathcal{L}^{t-1})=0$.*

Proof. Using $l_{\mathcal{A}}^{t-1}\geq\epsilon$ and $l_{\mathcal{L}}^{t-1}\geq\epsilon$, we get from (4.25) that

$$\begin{aligned}&\frac{\left(1-\delta_\omega^2\right)\beta\epsilon}{2}\sum_{t=1}^{\infty}\left\|\mathcal{A}^t-\mathcal{A}^{t-1}\right\|_F^2+\frac{(\gamma-1)\beta\epsilon}{4\gamma}\left(1-\delta_\omega^2\right)\sum_{t=1}^{\infty}\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2-\sum_{u=1}^{3}\frac{1}{2\beta_u^\infty}\left\|\mathcal{T}_u^\infty\right\|_F^2\\&\leq L\left(\mathcal{S}^\infty,\mathcal{A}^\infty,\mathcal{C}_u^\infty,\mathcal{L}^\infty;\mathcal{T}_u^\infty,\beta_u^\infty\right)+\frac{\left(1-\delta_\omega^2\right)\beta}{2}\sum_{t=1}^{\infty}l_{\mathcal{A}}^{t-1}\left\|\mathcal{A}^t-\mathcal{A}^{t-1}\right\|_F^2\\&\quad+\frac{(\gamma-1)\beta}{4\gamma}\left(1-\delta_\omega^2\right)\sum_{t=1}^{\infty}l_{\mathcal{L}}^{t-1}\left\|\mathcal{L}^t-\mathcal{L}^{t-1}\right\|_F^2\\&\leq L\left(\mathcal{S}^1,\mathcal{A}^1,\mathcal{C}_u^1,\mathcal{L}^1;\mathcal{T}_u^1,\beta_u^1\right)+\frac{\beta l_{\mathcal{A}}^0}{2}\left\|\mathcal{A}^1-\mathcal{A}^0\right\|_F^2+\frac{(\gamma-1)\beta l_{\mathcal{L}}^0}{4\gamma}\left\|\mathcal{L}^1-\mathcal{L}^0\right\|_F^2\\&\quad+\sum_{t=1}^{\infty}\sum_{u=1}^{3}\frac{\beta_u^t+\beta_u^{t+1}}{2\left(\beta_u^t\right)^2}\left\|\mathcal{T}_u^{t+1}-\mathcal{T}_u^t\right\|_F^2<\infty.\end{aligned}$$

Therefore, $\lim_{t\to\infty}(\mathcal{A}^t-\mathcal{A}^{t-1})=0$ and $\lim_{t\to\infty}(\mathcal{L}^t-\mathcal{L}^{t-1})=0$. ∎

Theorem 4.7. *Let $\{\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t,\mathcal{T}_u^t\}$ be a sequence generated by Algorithm 4.1. Suppose that the sequence $\{\mathcal{A}^t,\mathcal{L}^t\}_{t=1}^\infty$ is bound. Then any accumulation point of the sequence $\{\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t,\mathcal{T}_u^t\}$ is a Karush–Kuhn–Tucker (KKT) point of the optimization problem (4.2).*

Proof. If the sequence $\{\mathcal{A}^t,\mathcal{L}^t\}_{t=1}^\infty$ is bound, then by Lemmas 4.3 and 4.5, we know that the sequence $\{\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t,\mathcal{T}_u^t\}_{t=1}^\infty$ is bounded. From the Bolzano–Weierstrass theorem [4], Algorithm 4.1 has at least one accumulation point $\{\mathcal{S}^\star,\mathcal{A}^\star,\mathcal{C}_u^\star,\mathcal{L}^\star,\mathcal{T}_u^\star\}$ and there exists one sequence that converges to this accumulation point. Without loss of generality, we assume that the sequence is $\{\mathcal{S}^t,\mathcal{A}^t,\mathcal{C}_u^t,\mathcal{L}^t,\mathcal{T}_u^t\}_{t=1}^\infty$.

From the updating formula of $\mathcal{T}_u^{t+1}$, we have

$$\lim_{t\to\infty}\left\|\nabla_u\mathcal{A}^{t+1}-\mathcal{C}_u^{t+1}\right\|_F=\lim_{t\to\infty}\frac{1}{\beta_u^t}\left\|\mathcal{T}_u^{t+1}-\mathcal{T}_u^t\right\|_F=0. \tag{4.27}$$

Thus, $\lim_{t\to\infty}(\nabla_u\mathcal{A}^t-\mathcal{C}_u^t)=0$. Therefore we can obtain $\nabla_u\mathcal{A}^\star=\mathcal{C}_u^\star$ for $u\in[3]$. This indicates that this accumulation point can satisfy the feasible conditions of the GNBRL model.

For the $\mathcal{S}$-subproblem, by the first order optimal condition, we get

$$0\in\lambda_2\frac{\partial}{\partial\mathcal{S}}\Big|_{\mathcal{S}=\mathcal{S}^{t+1}}\|\mathcal{S}\|_{\ell_{F,1}^{\psi}}+\beta\nabla_{\mathcal{S}}f\left(\mathcal{A}^t,\mathcal{L}^t,\mathcal{S}^{t+1}\right). \tag{4.28}$$

By Lemma 4.6, letting $t\to\infty$ in (4.28), we have

$$0\in\lambda_2\frac{\partial}{\partial\mathcal{S}}\Big|_{\mathcal{S}=\mathcal{S}^{\star}}\|\mathcal{S}\|_{\ell_{F,1}^{\psi}}+\beta\nabla_{\mathcal{S}}f\left(\mathcal{A}^\star,\mathcal{L}^\star,\mathcal{S}^\star\right). \tag{4.29}$$

For the $\mathcal{A}$-subproblem, by the first order optimal condition, we get

$$\sum_{u=1}^{3}\left(\nabla_u^H\mathcal{T}_u^t+\beta_u^t\nabla_u^H\left(\nabla_u\mathcal{A}^{t+1}-\mathcal{C}_u^t\right)\right)+\beta\nabla_{\mathcal{A}}f\left(\hat{\mathcal{A}}^t,\mathcal{L}^t,\mathcal{S}^{t+1}\right)+\beta l_{\mathcal{A}}^t\left(\mathcal{A}^{t+1}-\hat{\mathcal{A}}^t\right)=0. \tag{4.30}$$

Combining $\lim_{t\to\infty}(\nabla_u\mathcal{A}^t-\mathcal{C}_u^t)=0$ and $\lim_{t\to\infty}(\mathcal{A}^t-\mathcal{A}^{t-1})=0$, one has $\lim_{t\to\infty}(\mathcal{C}_u^t-\mathcal{C}_u^{t-1})=0$. Thus

$$\lim_{t\to\infty}\mathcal{T}_u^t+\beta_u^t\left(\nabla_u\mathcal{A}^{t+1}-\mathcal{C}_u^t\right)=\lim_{t\to\infty}\mathcal{T}_u^t+\beta_u^t\left(\nabla_u\mathcal{A}^{t+1}-\mathcal{C}_u^{t+1}\right)=\lim_{t\to\infty}\mathcal{T}_u^{t+1}. \tag{4.31}$$

According to the definition and the boundedness of $\mathcal{L}^t$, we obtain that $l_{\mathcal{A}}^t$ and $\omega_{\mathcal{A}}^t$ are bound. Combining this with Lemma 4.6 gives

$$\begin{aligned}\lim_{t\to\infty}\nabla_{\mathcal{A}}f\left(\hat{\mathcal{A}}^t,\mathcal{L}^t,\mathcal{S}^{t+1}\right)&=\lim_{t\to\infty}\nabla_{\mathcal{A}}f\left(\mathcal{A}^t+\omega_{\mathcal{A}}^t\left(\mathcal{A}^t-\mathcal{A}^{t-1}\right),\mathcal{L}^t,\mathcal{S}^{t+1}\right)\\&=\lim_{t\to\infty}\nabla_{\mathcal{A}}f\left(\mathcal{A}^{t+1},\mathcal{L}^{t+1},\mathcal{S}^{t+1}\right),\\\lim_{t\to\infty}l_{\mathcal{A}}^t\left(\mathcal{A}^{t+1}-\hat{\mathcal{A}}^t\right)&=\lim_{t\to\infty}l_{\mathcal{A}}^t\left(\mathcal{A}^{t+1}-\mathcal{A}^t-\omega_{\mathcal{A}}^t\left(\mathcal{A}^t-\mathcal{A}^{t-1}\right)\right)=0.\end{aligned} \tag{4.32}$$

From (4.30), (4.31), and (4.32), we can easily observe that

$$\sum_{u=1}^{3}\nabla_u^H\mathcal{T}_u^\star+\beta\nabla_{\mathcal{A}}f\left(\mathcal{A}^\star,\mathcal{L}^\star,\mathcal{S}^\star\right)=0. \tag{4.33}$$

For the $\mathcal{C}_u$-subproblem, by the first order optimal condition, we get

$$\begin{aligned}0&\in\alpha_u\frac{\partial}{\partial\mathcal{C}_u}\Big|_{\mathcal{C}_u=\mathcal{C}_u^{t+1}}\|\mathcal{C}_u\|_{\psi}-\mathcal{T}_u^t+\beta_u^t\left(C_u^{t+1}-\nabla_u\mathcal{A}^{t+1}\right)\\&=\alpha_u\frac{\partial}{\partial\mathcal{C}_u}\Big|_{\mathcal{C}_u=\mathcal{C}_u^{t+1}}\|\mathcal{C}_u\|_{\psi}-\mathcal{T}_u^{t+1}.\end{aligned} \tag{4.34}$$

Letting $t\to\infty$, we have

$$0\in\alpha_u\frac{\partial}{\partial\mathcal{C}_u}\Big|_{\mathcal{C}_u=\mathcal{C}_u^{\star}}\|\mathcal{C}_u\|_{\psi}-\mathcal{T}_u^\star. \tag{4.35}$$

For the $\mathcal{L}$-subproblem, by the first order optimal condition, we get

$$0 \in \lambda_1 \frac{\partial}{\partial \mathcal{L}}\Big|_{\mathcal{L}=\mathcal{L}^{t+1}} \|\mathcal{L}\|_\psi + \beta \nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \hat{\mathcal{L}}^t, \mathcal{S}^{t+1}\right) + \beta l_{\mathcal{L}}^t \left(\mathcal{L}^{t+1} - \hat{\mathcal{L}}^t\right). \tag{4.36}$$

According to the boundedness of $\mathcal{A}^{t+1}$, we obtain that $l_{\mathcal{L}}^t$ is bound. Combining this with Lemma 4.6 gives

$$\begin{aligned} \lim_{t\to\infty} \nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \hat{\mathcal{L}}^t, \mathcal{S}^{t+1}\right) &= \lim_{t\to\infty} \nabla_{\mathcal{L}} f\left(\mathcal{A}^{t+1}, \mathcal{L}^{t+1}, \mathcal{S}^{t+1}\right), \\ \lim_{t\to\infty} l_{\mathcal{L}}^t \left(\mathcal{L}^{t+1} - \hat{\mathcal{L}}^t\right) &= 0. \end{aligned} \tag{4.37}$$

From (4.36) and (4.37), we can easily observe that

$$0 \in \lambda_1 \frac{\partial}{\partial \mathcal{L}}\Big|_{\mathcal{L}=\mathcal{L}^\star} \|\mathcal{L}\|_\psi + \beta \nabla_{\mathcal{L}} f\left(\mathcal{A}^\star, \mathcal{L}^\star, \mathcal{S}^\star\right). \tag{4.38}$$

So, by $\nabla_u \mathcal{A}^\star = \mathcal{C}_u^\star$ for $u \in [3]$, (4.29), (4.33), (4.35), and (4.38), we have that $\{\mathcal{S}^\star, \mathcal{A}^\star, \mathcal{C}_u^\star, \mathcal{L}^\star, \mathcal{T}_u^\star\}$ is a KKT point of (4.2). ∎

4.3. A stopping criterion for the GNBRL algorithm. To measure the precision of the optimal solution obtained by Algorithm 4.1, we used the relative KKT residual $\vartheta = \max\{\vartheta_1, \vartheta_2\}$. Here $\vartheta_1 = \max\{\vartheta_S, \vartheta_A, \vartheta_{C_1}, \vartheta_{C_2}, \vartheta_{C_3}, \vartheta_L\}$ with

$$\vartheta_S = \frac{\left\|\mathcal{S} - \operatorname{prox}_{\lambda_2\|\cdot\|_{\ell_{F,1}^\psi}}\left(\mathcal{S} - \beta\nabla_{\mathcal{S}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\right)\right\|_F}{1 + \|\mathcal{S}\|_F + \|\beta\nabla_{\mathcal{S}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\|_F}, \quad \vartheta_{C_u} = \frac{\left\|\mathcal{C}_u - \operatorname{prox}_{\alpha_u\|\cdot\|_\psi}\left(\mathcal{C}_u + \mathcal{T}_u\right)\right\|_F}{1 + \|\mathcal{C}_u\|_F + \|\mathcal{T}_u\|_F},$$

$$\vartheta_A = \frac{\left\|\sum_{u=1}^3 \nabla_u^H \mathcal{T}_u + \beta\nabla_{\mathcal{A}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\right\|_F}{1 + \sum_{u=1}^3 \|\nabla_u^H \mathcal{T}_u\|_F + \|\beta\nabla_{\mathcal{A}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\|_F}, \quad \vartheta_L = \frac{\left\|\mathcal{L} - \operatorname{prox}_{\lambda_1\|\cdot\|_\psi}\left(\mathcal{L} - \beta\nabla_{\mathcal{L}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\right)\right\|_F}{1 + \|\mathcal{L}\|_F + \|\beta\nabla_{\mathcal{L}} f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)\|_F},$$

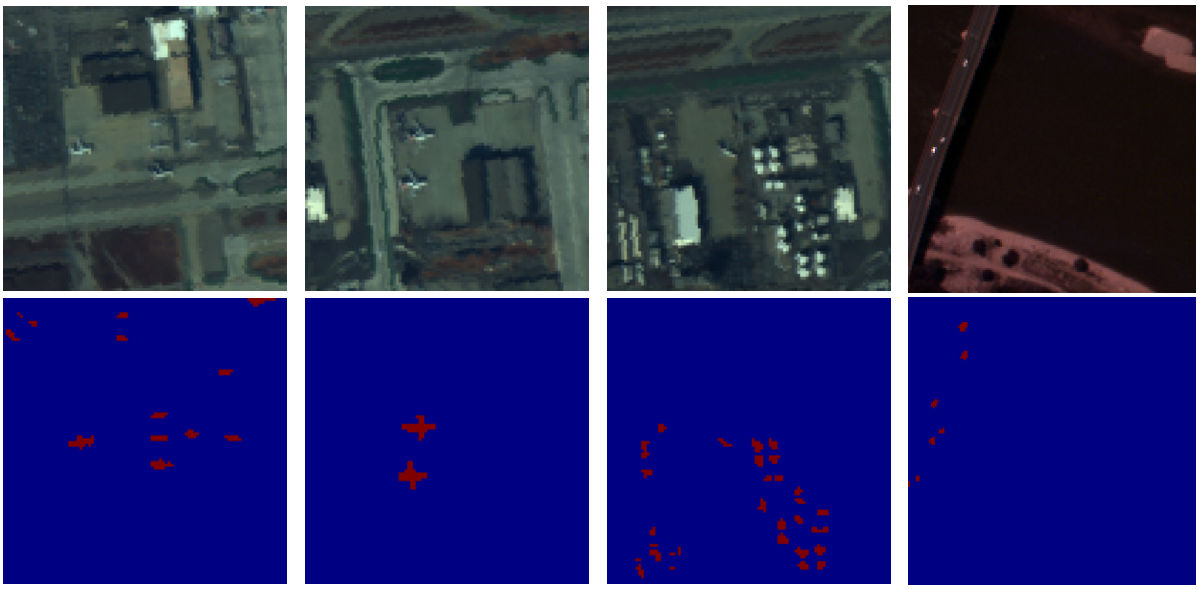
and $\vartheta_2 = \max\{\chi_1, \chi_2, \chi_3\}$ with

$$\chi_u = \frac{\|\nabla_u \mathcal{A} - \mathcal{C}_u\|_F}{1 + \|\nabla_u \mathcal{A}\|_F + \|\mathcal{C}_u\|_F}.$$

For a given tolerance ε, we will terminate the Algorithm 4.1 when $\vartheta < \varepsilon$ or the maximum number of iterations is reached.

5. Experiments. In this section, a series of experiments are conducted to evaluate the performance of the proposed algorithm on a server with 16 logical CPU cores and 16 GB memory. We implemented the codes of all algorithms in MATLAB 2022a and did not apply any preprocessing to ensure fairness.

For comparison, we use some general methods in the experiments, such as classic global RX [40], four matrix based methods (namely, RPCA [45], LRASR [53], LSMAD [62], GTVLRR [9]), and three tensor based methods (namely, PTA [28], TPCA [8], TLRSR [49]). Besides the visual observation of the resulted anomaly maps, the experimental results are evaluated by the receiver operating characteristic (ROC) curve [23] and the area under the ROC curve



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 2. *Pseudo-color images and ground truth maps of the four HSI data sets.*

(AUC) [24] to quantitatively assess the detection accuracy. The ROC curve plots the varying relationship of the probability of detection (PD) and false alarm rate for various possible thresholds. The area under this curve is calculated as the AUC, and a higher AUC score indicates a better detection method. The Algorithm 4.1 is terminated when $\vartheta < \varepsilon = 10^{-3}$ or the maximum number of 300 iterations is reached.

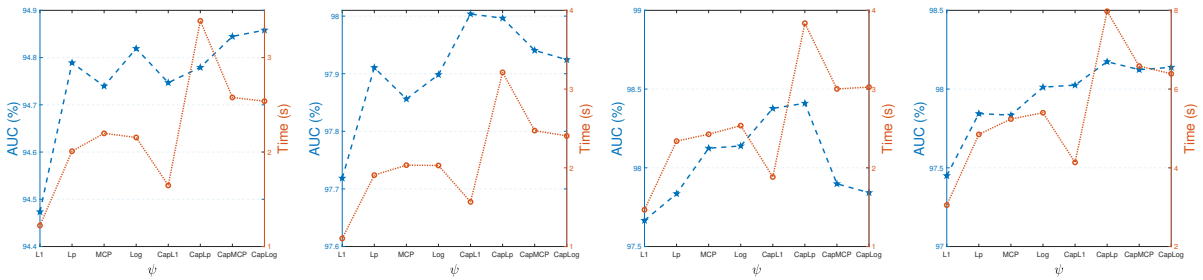
5.1. Hyperspectral data. The airport-beach-urban[1] [22] scenes are employed to evaluate the effectiveness of the proposed method for HAD tasks. The information from the data sets is as follows.

(1) Airport: The first data set was captured by the airborne visible/infrared imaging spectrometer (AVIRIS) [59] sensor over Los Angeles, which has a spatial resolution of 7.1 m. This data set has a spatial dimension of 100 by 100 and 205 spectral bands. The false color image and the ground truth map are shown in Figures 2(a)–(b).

(2) Urban: The second data set was generated based on an AVIRIS data set over Los Angeles with a spatial resolution of 7.1 m. This data set has a spatial dimension of 205 by 100 and 100 spectral bands. The image scene and the ground truth map are illustrated in Figure 2(c).

(3) Beach: The third data set was captured by the ROSIS-03 sensor in Pavia, which has a spatial resolution of 1.3 m. This data set has a spatial dimension of 102 by 150 and 150 spectral bands. A sample band of the scene, together with its corresponding ground truth, is shown in Figure 2(d).

[1] http://xudongkang.weebly.com/data-sets.html



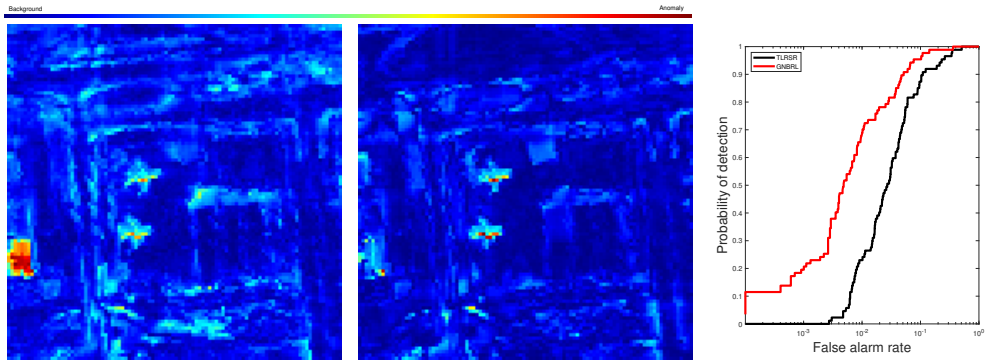
(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 3. *AUC values (%) and corresponding running times of GNBRL with different nonconvex functions ψ for each data set.*

5.2. Performance of different nonconvex functions. In this subsection, the performance of our proposed method GNBRL for HAD under different nonconvex functions is investigated. We compare eight functions, namely "L1," "Lp," "MCP," "Log," "CapL1," "CapLp," "CapMCP," "CapLog." Figure 3 shows the curves of the AUC values (%) and the corresponding running times for different nonconvex functions ψ. We can see from the figure that the highest AUC values correspond to different functions ψ under different data sets. We pursue a general surrogate for approximating low rank and sparsity, which can be flexibly chosen according to different scenarios. This allows us to better capture the low rank structure of the background and the sparsity of the anomaly. We can observe that the convex function L1 performs worse than all nonconvex functions in terms of AUC values. We can also observe that the non-Cap-type functions generally perform worse than the corresponding Cap-type functions in terms of AUC values. Among the Cap-type functions, CapL1 and CapLp achieve relatively stable and high AUC values across different data sets, but CapLp requires more running time. Therefore, for simplicity, we use the CapL1 function in the following experiments.

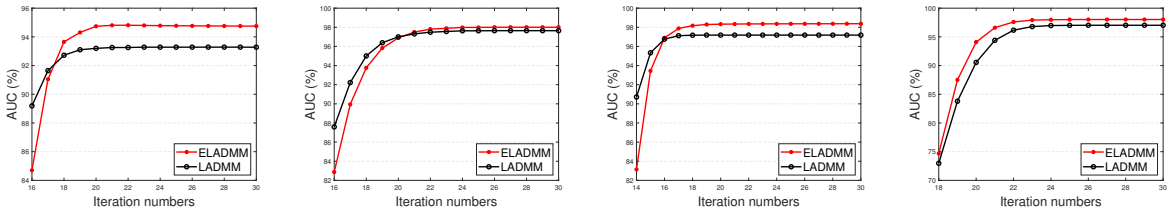
5.3. Effects of the dictionary constraint. In this subsection, we evaluate the effectiveness of dictionary constraint by comparing the restoration results obtained by TLRSR and GNBRL. Figure 4 shows the detection maps and ROC curves obtained by TLRSR and GNBRL for the "Airport2" data set. As shown in Figures 4(a)–(b), the proposed GNBRL cannot only provide clearer maps for the anomalies than its counterparts, but also has lower test statistics for background pixels. Figure 4(c) reports the ROC curves by TLRSR and GNBRL. It can be observed that GNBRL achieves a higher detection probability than TLRSR for any false alarm rate. This further validates the effectiveness of the dictionary constraint.

5.4. Effects of extrapolation strategy. In this subsection, we demonstrate the effectiveness of the extrapolation strategy through some illustrative examples. Figure 5 shows the AUC values (%) versus the number of iterations for LADMM (i.e., $\omega_{\mathcal{A}}^t = \omega_{\mathcal{L}}^t = 0$ in ELADMM) and ELADMM. In order to better observe the difference between them, we only display the index transformation trends after about 16 iterations. From the figure, we can see that ELADMM always achieves better performance with higher AUC values. Therefore, in the next section, we use ELADMM to solve GNBRL and compare it with other methods.



(a) TLRSR (b) GNBRL (c) ROC curves

Figure 4. *Detection maps and ROC curves obtained by TLRSR and GNBRL for the Airport2 data set, respectively.*



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

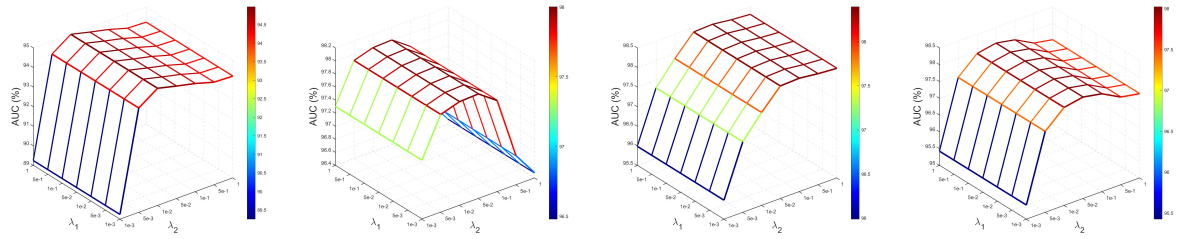
Figure 5. *AUC values (%) with respect to the iteration numbers for LADMM and ELADMM.*

5.5. Parameters setting. For the proposed GNBRL model, six parameters α_1, α_2, α_3, λ_1, λ_2, β affect the performance. We choose $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$ in all experiments for simplicity since the detection performance is stable as long as $\alpha_1\alpha_2\alpha_3 \neq 0$.

The performance is first compared for different values of λ_1 and λ_2. The values of λ_1 and λ_2 are set from $\{1e\text{-}3, 5e\text{-}3, 1e\text{-}2, 5e\text{-}2, 1e\text{-}1, 5e\text{-}1, 1\}$. Figure 6 shows the surfaces of AUC values in terms of different λ_1's and λ_2's. It can be observed that the change in λ_1 has little effect on GNBRL. It can also be observed that the GNBRL can work stably when $\lambda_1 \in [1e\text{-}3, 1]$ and $\lambda_2 \in [5e\text{-}2, 1]$ except for Airport2. Therefore, for simplicity, we choose $\lambda_1 = 5e\text{-}1$ and $\lambda_2 = 5e\text{-}2$ in all experiments.

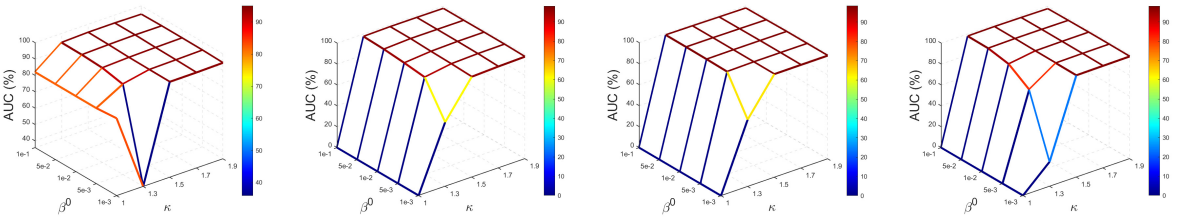
Second, the performance of GNBRL for different β's is investigated. For simplicity, we set $\beta^{t+1} = \min\{\kappa\beta^t, 1e8\}$. We choose β^0 and κ from the sets of $\{1e\text{-}3, 5e\text{-}3, 1e\text{-}2, 5e\text{-}2, 1e\text{-}1\}$ and $\{1, 1.3, 1.5, 1.7, 1.9\}$, respectively. For fairness, when $\kappa = 1$, we choose β^0 from the set $\{1e2, 1e4, 1e6, 1e8, 1e10\}$. The AUC values surfaces versus different β^0's and κ's are shown in Figure 7. As can be seen, $\kappa > 1$ basically obtained a higher AUC value than $\kappa = 1$. Additionally, changing β^0 has little effect on GNBRL. Therefore, for simplicity, we set $\beta^{t+1} = \min\{1.5\beta^t, 1e8\}$ with $\beta^0 = 5e\text{-}2$ in all experiments.

5.6. Numerical convergence. The numerical convergence is investigated on four data sets. Figure 8 shows an empirical analysis of the convergence of CF2-GNBRL. It can be



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 6. *Surfaces of AUC values (%) with different* λ_1*'s and* λ_2*'s.*



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 7. *Surfaces of AUC values (%) with different* β^0*'s and* κ*'s.*

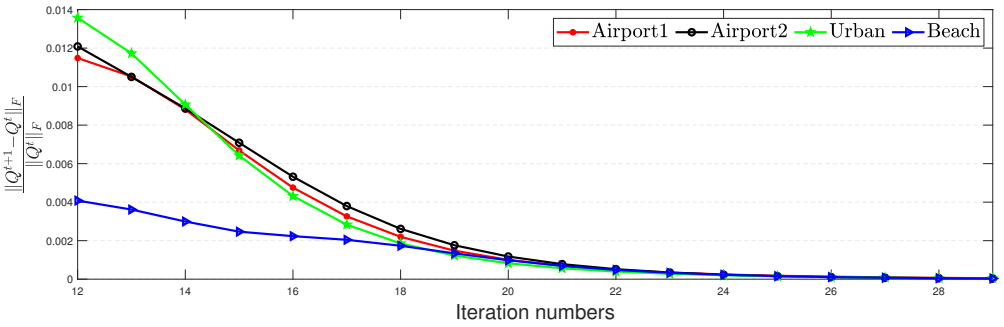


Figure 8. *Relative change of* $Q = [\mathcal{A}(:); \mathcal{L}(:); \mathcal{S}(:)]$ *with respect to the iteration numbers for CF2-GNBRL.*

observed that as the number of iterations increases, the curves of the relative changes in Q monotonically decrease, and the iteration stabilizes after 24 iterations. The convergence of CF2-GNBRL is numerically guaranteed.

5.7. Detection performance. We evaluate and compare the detection performance of our proposed GNBRL and CF2-GNBRL with eight other state-of-the-art detectors. Figure 9 shows the two-dimensional detection results of these methods. The results show that GNBRL and CF2-GNBRL can effectively detect the anomalies, and CF2-GNBRL has a stronger background suppression effect than GNBRL. It also shows that the responses of abnormal pixels in the detection maps from RX, RPCA, LRASR, LSMAD, and GTVLRR are lower than those from GNBRL and CF2-GNBRL. For the TPCA method, the contrast between the

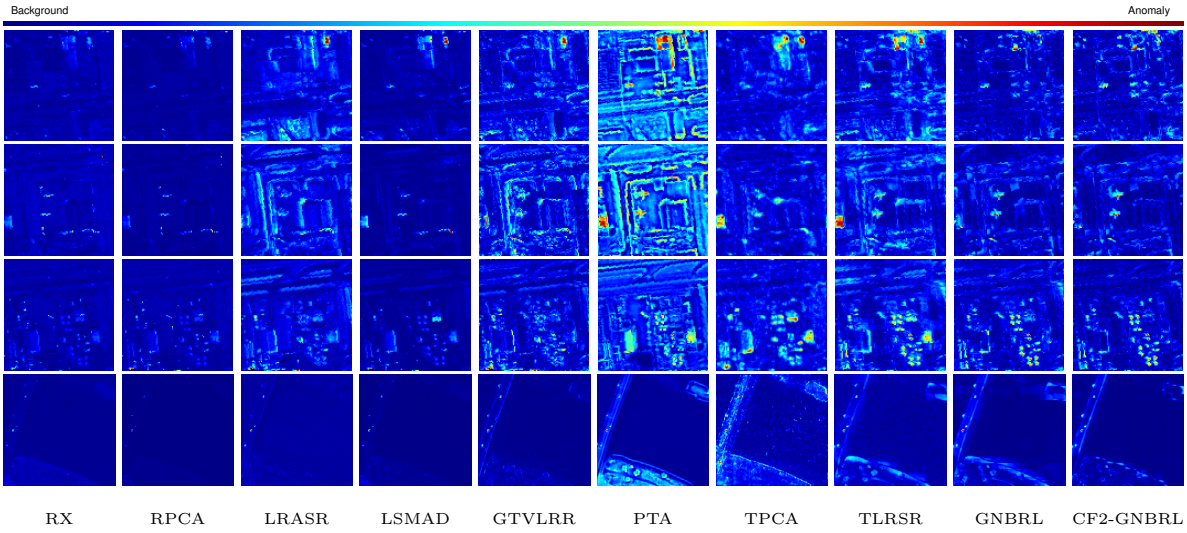
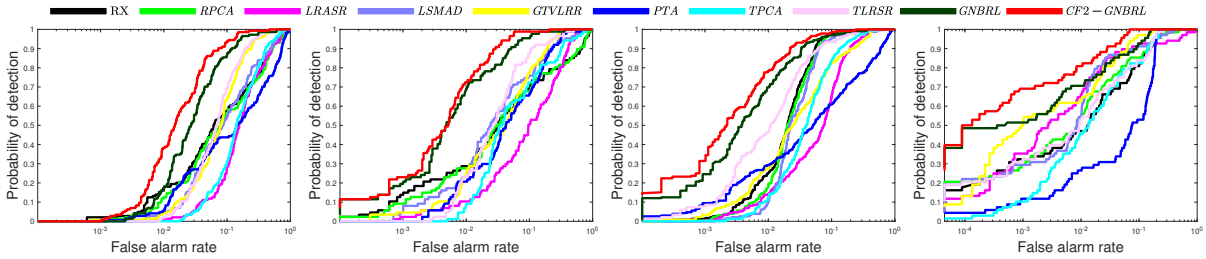


Figure 9. *Target detection results by different methods for the four data sets.*



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 10. *ROC curves obtained by different methods for the four data sets.*

background and the anomaly is not clear as this model does not compress the background well. For PTA and TLRSR methods, the detection maps are unsatisfactory due to the complex visualization of the background region, especially for PTA. Overall, these results demonstrate the superiority of our proposed method.

Additionally, to quantitatively compare the performance of the proposed methods, the ROC curves and AUC values (%) of different methods on the four data sets are shown in Table 1 and Figure 10, respectively. The best results for each dataset in Table 1 are highlighted in bold. Figure 10 shows that the GNBRL method has a higher detection probability than other methods we compare with for any false alarm rate. Moreover, the CF2-GNBRL method further improves the detection probability compared with the GNBRL method. The detection probabilities of the GNBRL and CF2-GNBRL methods are approximately 1 when the false alarm rate is about 10^{-1}. Furthermore, the detection probability of the GNBRL and CF2-GNBRL methods is much higher than that of other methods when the false alarm rate is around 10^{-2}. Table 1 shows that our proposed method GNBRL achieves higher AUC values than all the other methods we compare with. By applying the CF2 strategy, the CF2-GNBRL

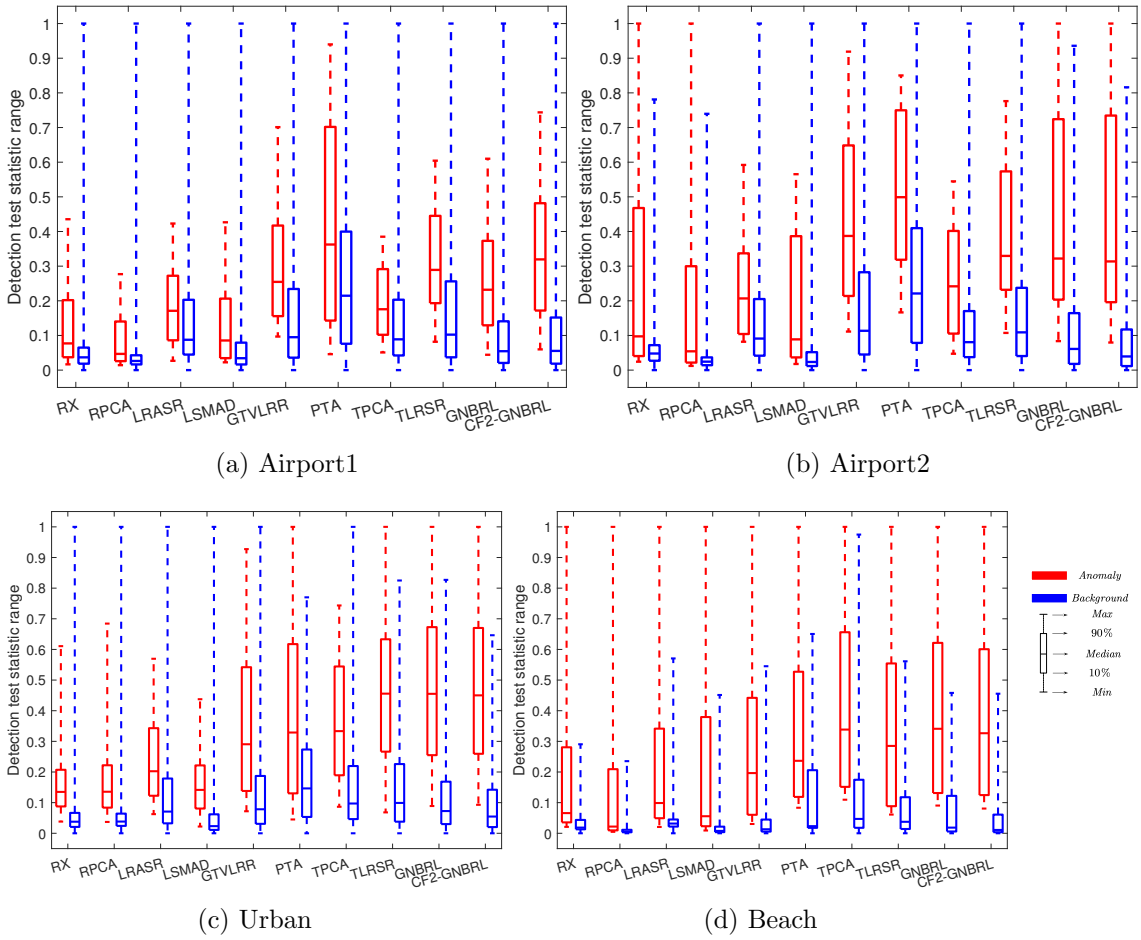
Table 1
Comparison of AUC values (%) and running time (s) of different methods for the four data sets.

HSI	Airport1		Airport2		Urban		Beach	
Algorithm	AUC (%)	Time (s)	AUC (%)	Time (s)	AUC (%)	Time (s)	AUC (%)	Time (s)
RX	82.21	0.42	84.03	0.41	96.92	0.41	95.39	0.04
RPCA	80.89	8.00	84.31	7.44	96.58	6.98	95.99	1.95
LRASR	77.28	53.81	86.48	70.13	92.89	47.51	95.65	104.90
LSMAD	83.39	9.54	92.17	8.60	96.05	8.74	97.06	7.65
GTVLRR	90.04	171.47	88.89	227.16	93.73	229.16	98.02	378.60
PTA	73.30	13.50	90.95	20.96	82.57	24.89	90.61	29.11
TPCA	80.22	30.91	88.90	30.62	93.69	22.15	95.82	21.71
TLRSR	90.56	3.44	94.57	3.63	97.10	3.58	95.98	5.84
GNBRL	94.75	1.60	98.00	1.50	98.38	1.91	98.03	4.01
CF2-GNBRL	**96.84**	27.14	**98.81**	31.63	**98.98**	31.40	**99.24**	83.06

method further increases the AUC values, surpassing the performance of the GNBRL method. The AUC values obtained by CF2-GNBRL on the four data sets are 6.93%, 4.48%, 1.94%, and 3.40% higher than TLRSR, respectively, which is the best method among all the methods except our proposed GNBRL and CF2-GNBRL methods. In conclusion, the CF2-GNBRL method is far superior than other methods. In addition, we report the running time of each algorithm in Table 1. It can be observed that from the results on the Beach dataset, the proposed GNBRL uses slightly more time compared to the RX and the RPCA methods, while using less time than other methods. For the other three datasets, the proposed GNBRL is the second fastest method. It is about twice as fast as the third fastest method TLRSR. However, CF2-GNBRL consumes much more time than GNBRL because it requires processing multiple smaller subtensors. Nevertheless, this trade-off is worthwhile for achieving better performance in most cases. In summary, we can choose GNBRL if we prioritize speed; otherwise, we can opt for CF2-GNBRL.

To further evaluate the performance of GNBRL and CF2-GNBRL, we use the normalized background-anomaly separation maps to show the abilities of different methods for separating anomaly pixels from background pixels. The larger the distance between the anomalous and background boxes, the better the separation performance. As shown in Figure 11, CF2-GNBRL reaches the best performance because it has the largest distance between the anomalous box and the background box. The background-anomaly separation of GNBRL is close to that of CF2-GNBRL, which achieves the second best performance. The conclusions obtained by Figure 11 are also consistent with those obtained by Figures 9 and 10 and Table 1, and all these demonstrate that both our GNBRL method and the CF2 strategy are effective in improving the detection performance.

6. Conclusions and future work. In this paper, we have proposed a novel GNBRL model for HAD. In our model, through 3DCTV regularization, the background representation learning model with dictionary constraint is adopted. Then, a class of nonconvex functions is employed on low rank and sparse terms, respectively. An error bound of the GNBRL model is established. Moreover, an ELADMM algorithm is developed to solve the GNBRL model



(a) Airport1 (b) Airport2

(c) Urban (d) Beach

Figure 11. *Separability maps of different methods for the four data sets.*

and its convergence is established. Finally, a general and easily implementable coarse to fine framework called CF2 is designed, which can effectively boost the performance of the GNBRL model. The experimental results on different HSI datasets demonstrate the superior performance of our proposed methods.

The CF2-GNBRL model is a robust and accurate anomaly detection method for various HSI tasks and data, such as target detection, material identification, and environmental monitoring. In the future, it would be of great interest to improve the efficiency of CF2 strategy. Although the CF2 strategy is helpful in improving performance, it requires solving the model many times, which results in a significant reduction in efficiency. Therefore, developing an efficient method similar to the CF2 strategy is necessary.

Appendix A. Preliminaries of tensors.

In Appendix A, we give some definitions of tensors used throughout this paper.

Definition A.1 (f-diagonal tensor [25]). *A tensor is called f-diagonal if each of its corresponding frontal slices is a diagonal matrix.*

Definition A.2 (conjugate transpose [25]). *The conjugate transpose of a tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is the tensor* $\mathcal{X}^H \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ *obtained by conjugate transposing each of the corresponding frontal slices and then reversing the order of the resulting transposed frontal slices* 2 *through* n_3.

Definition A.3 (identity tensor [25]). *The identity tensor* $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ *is the tensor whose first frontal slice is the* $n \times n$ *identity matrix, and the remaining frontal slices are all zeros.*

Definition A.4 (orthogonal tensor [25]). *An orthogonal tensor* $\mathcal{X} \in \mathbb{R}^{n \times n \times n_3}$ *is a tensor that satisfies the following condition:* $\mathcal{X}^H * \mathcal{X} = \mathcal{X} * \mathcal{X}^H = \mathcal{I}$.

Appendix B. Proximal mapping. The proximal operator of z with respect to ψ is defined as

$$\operatorname{prox}_{\lambda\psi}(z) := \arg\min_x \ \lambda\psi(x) + \frac{1}{2}(x-z)^2.$$

For the specific forms of the nonconvex functions ψ stated in Theorem 3.2, the proximal mappings often have analytical expressions, which we summarize as follows:

- L1: the proximal mapping of $\psi^{\mathrm{L1}}(x)$ is given by [46]
$$\operatorname{prox}_{\lambda\psi^{\mathrm{L1}}}(z) := \operatorname{sign}(z)\max\{|z|-\lambda, 0\}.$$
- Lp: for any $0 < p < 1$, the proximal mapping of $\psi^{\mathrm{Lp}}(x)$ is given by [32]
$$\operatorname{prox}_{\lambda\psi^{\mathrm{Lp}}}(z) := \begin{cases} 0 & \text{if } |z| < \pi_2, \\ \{0, \operatorname{sgn}(z)\pi_1\} & \text{if } |z| = \pi_2, \\ \operatorname{sgn}(z)\pi_\star & \text{if } |z| > \pi_2, \end{cases}$$
where $\pi_1 = (2\lambda(1-p))^{\frac{1}{2-p}}$, $\pi_2 = \pi_1 + \lambda p \pi_1^{p-1}$, and $\pi_\star \in (\pi_1, |z|)$ is the solution of $g(\pi) = \pi + \lambda p \pi^{p-1} - |z| = 0$ with $\pi > 0$.
- MCP: for any $\alpha > 0$, the proximal mapping of $\psi^{\mathrm{MCP}}(x)$ is given by [60]
$$\operatorname{prox}_{\lambda\psi^{\mathrm{MCP}}}(z) := \begin{cases} 0 & \text{if } |z| \leqslant \lambda, \\ \frac{\operatorname{sign}(z)(|z|-\lambda)}{1-\lambda/\alpha} & \text{if } \lambda < |z| \leqslant \alpha, \\ z & \text{if } |z| > \alpha. \end{cases}$$
- Logarithm: for any $\theta > 0$, the proximal mapping of $\psi^{\mathrm{Log}}(x)$ is given by [14]
$$\operatorname{prox}_{\lambda\psi^{\mathrm{Log}}}(z) := \operatorname{sign}(z) y,$$
where y is an optimal solution of the problem $y = \arg\min_{x \in \mathfrak{C}}\{\lambda\psi^{\mathrm{Log}}(x) + \frac{1}{2}(x-z)^2\}$, and $\mathfrak{C}$ is a set composed of 3 elements or 1 element. If $a := (|z|-\theta)^2 - 4(\lambda - \theta|z|) \geqslant 0$, $\mathfrak{C} = \{0, \max\{\frac{|z|-\theta+\sqrt{a}}{2}, 0\}, \max\{\frac{|z|-\theta-\sqrt{a}}{2}, 0\}\}$, otherwise, $\mathfrak{C} = \{0\}$.
- Capped folded functions: for $\psi^{\mathrm{Cap}}(x) := \begin{cases} c\psi(x) & \text{if} 0 \leq x < v, \\ c_0 & \text{if} x \geq v \end{cases}$ with $c_0, c > 0$, and $c\psi(v) = c_0$. The proximal mapping of $\psi^{\mathrm{Cap}}(x)$ is given by [36]
$$\operatorname{prox}_{\lambda\psi^{\mathrm{Cap}}}(z) := \begin{cases} u_1 & \text{if } \psi^{\mathrm{Cap}}(u_1) + \frac{1}{2}(u_1 - z)^2 \leq \psi^{\mathrm{Cap}}(u_2) + \frac{1}{2}(u_2 - z)^2, \\ u_2 & \text{otherwise}. \end{cases}$$
Here $u_1 = \min\{\max\{\operatorname{prox}_{c\lambda\psi}(z), 0\}, v\}$ and $u_2 = \max\{z, v\}$.

Lemma B.1 (see [38]). *Let us consider a tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which has t-SVD $\mathcal{Z} = \mathcal{U} * \mathcal{F} * \mathcal{V}^H$. Then the solutions $\mathcal{X}^\star$ to the problem*

$$\min_{\mathcal{X}} \ \lambda \|\mathcal{X}\|_\psi + \frac{1}{2}\|\mathcal{X} - \mathcal{Z}\|_F^2 \tag{B.1}$$

are

$$\mathcal{X}^\star = \text{prox}_{\lambda\|\cdot\|_\psi}(\mathcal{Z}) := \mathcal{U} * \mathcal{D} * \mathcal{V}^H,$$

where $\mathcal{D}$ is an f-diagonal tensor and $\bar{D}_{i,i}^{(k)} = \text{prox}_{\lambda\psi}(\bar{F}_{i,i}^{(k)})$.

Appendix C. Proof of Theorem 3.9.

Before we prove Theorem 3.9, we first present some lemmas.

Lemma C.1. *Concavity of $\psi(x)$ and $\psi(0) = 0$ means its subadditivite.*

Proof. For $x_1 \geq 0$ and $x_2 \geq 0$, concavity implies

$$\psi(x_1) = \psi\left(\frac{x_1}{x_1+x_2}(x_1+x_2) + \frac{x_2}{x_1+x_2}\cdot 0\right) \geq \frac{x_1}{x_1+x_2}\psi(x_1+x_2) + \frac{x_2}{x_1+x_2}\psi(0)$$

and

$$\psi(x_2) = \psi\left(\frac{x_2}{x_1+x_2}(x_1+x_2) + \frac{x_1}{x_1+x_2}\cdot 0\right) \geq \frac{x_2}{x_1+x_2}\psi(x_1+x_2) + \frac{x_1}{x_1+x_2}\psi(0).$$

Then,

$$\psi(x_1) + \psi(x_2) \geq \psi(x_1+x_2) + \psi(0) = \psi(x_1+x_2). \qquad \blacksquare$$

Lemma C.2 (see [41, Theorem 1]). *Suppose that B and S are two same-sized matrices and that ψ satisfies Assumption 3.1. Then $\|B+S\|_\psi \leq \|B\|_\psi + \|S\|_\psi$.*

Then we proceed to prove Theorem 3.9.

Proof. (1) Denote $\mathcal{L} = \mathcal{B} - \mathcal{S}$. By Lemma C.2, we obtain that

$$\|\mathcal{B}\|_\psi = \|\mathcal{L}+\mathcal{S}\|_\psi = \sum_{k=1}^{n_3}\frac{1}{n_3}\left\|\bar{L}^{(k)} + \bar{S}^{(k)}\right\|_\psi \leq \sum_{k=1}^{n_3}\frac{1}{n_3}\left\|\bar{L}^{(k)}\right\|_\psi + \sum_{k=1}^{n_3}\frac{1}{n_3}\left\|\bar{S}^{(k)}\right\|_\psi = \|\mathcal{L}\|_\psi + \|\mathcal{S}\|_\psi,$$

which implies that $\|\mathcal{B}-\mathcal{S}\|_\psi \geq \|\mathcal{B}\|_\psi - \|\mathcal{S}\|_\psi$. Thus, we complete the proof of statement (1).

(2) According to the definition of $\|\cdot\|_{\ell_{F,1}^\psi}$, we have

$$\begin{aligned}\|\mathcal{B}-\mathcal{S}\|_{\ell_{F,1}^\psi} &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:) - \mathcal{S}(i,j,:)\|_F\right) \leq \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:)\|_F + \|\mathcal{S}(i,j,:)\|_F\right)\\ &\leq \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:)\|_F\right) + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{S}(i,j,:)\|_F\right) = \|\mathcal{B}\|_{\ell_{F,1}^\psi} + \|\mathcal{S}\|_{\ell_{F,1}^\psi},\end{aligned}$$

where the first inequality follows from the triangle inequality of $\|\cdot\|_F$, the second inequality follows from the fact that ψ is a subadditive function, as shown by using Lemma C.1. Hence, we establish result (2) in this lemma.

(3) From the definition of $\|\cdot\|_{\ell^{\psi}_{F,1}}$, one has

$$\|\mathcal{B}\|_{\ell^{\psi}_{F,1}} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:)\|_F\right) \geq \psi\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\|\mathcal{B}(i,j,:)\|_F\right) \geq \psi\left(\|\mathcal{B}\|_F\right) \tag{C.1}$$

and

$$\|\mathcal{B}\|_{\ell^{\psi}_{F,1}} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:)\|_F\right) \leq \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{B}(i,j,:)\|_1\right) \leq \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\psi\left(|\mathcal{B}_{ijk}|\right) = \|\mathcal{B}\|_{\psi,1}, \tag{C.2}$$

where the first inequality of (C.1) and the second inequality of (C.2) come from the fact that ψ is a subadditive function, a result of Lemma C.1, and the last inequality of (C.1) is based on the triangle inequality of $\|\cdot\|_F$. ∎

REFERENCES

[1] W. Bian and X. Chen, *A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty*, SIAM J. Numer. Anal., 58 (2020), pp. 858–883, https://doi.org/10.1137/18m1186009.

[2] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, *Simultaneous analysis of Lasso and Dantzig selector*, Ann. Statist., 37 (2009), pp. 1705–1732, https://doi.org/10.1214/08-aos620.

[3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, *Hyperspectral remote sensing data analysis and future challenges*, IEEE Geosci. Remote Sens. Mag., 1 (2013), pp. 6–36, https://doi.org/10.1109/mgrs.2013.2244672.

[4] A. Browder, *Mathematical Analysis*, Undergrad. Texts Math., Springer New York, 2012, https://doi.org/10.1007/978-1-4612-0715-3.

[5] P. S. Bullen, D. S. Mitrinović, and P. M. Vasić, *The Power Means*, Springer, Dordrecht, Netherlands, 1988, pp. 132–214, https://doi.org/10.1007/978-94-017-2226-1_3.

[6] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb, *Blind image fusion for hyperspectral imaging with the directional total variation*, Inverse Problems, 34 (2018), 044003, https://doi.org/10.1088/1361-6420/aaaf63.

[7] X. Chen, F. Xu, and Y. Ye, *Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization*, SIAM J. Sci. Comput., 32 (2010), pp. 2832–2852, https://doi.org/10.1137/090761471.

[8] Z. Chen, B. Yang, and B. Wang, *A preprocessing method for hyperspectral target detection based on tensor principal component analysis*, Remote Sens., 10 (2018), 1033, https://doi.org/10.3390/rs10071033.

[9] T. Cheng and B. Wang, *Graph and total variation regularized low-rank representation for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 58 (2020), pp. 391–406, https://doi.org/10.1109/tgrs.2019.2936609.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3-D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095, https://doi.org/10.1109/tip.2007.901238.

[11] A. Danielyan, V. Katkovnik, and K. Egiazarian, *BM3D frames and variational image deblurring*, IEEE Trans. Image Process., 21 (2012), pp. 1715–1728, https://doi.org/10.1109/tip.2011.2176954.

[12] J. H. Friedman, *Fast sparse regression and classification*, Int. J. Forecast., 28 (2012), pp. 722–738, https://doi.org/10.1016/j.ijforecast.2012.05.001.

[13] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, *BS^3LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 61 (2023), 5504218, https://doi.org/10.1109/tgrs.2023.3246565.

[14] P. Gong, C. Zhang, Z. Lu, J. Z. Huang, and J. Ye, *A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems*, JMLR Workshop Conf. Proc., 28 (2013), pp. 37–45.

[15] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, *Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery*, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 7 (2014), pp. 2351–2366, https://doi.org/10.1109/jstars.2014.2302446.

[16] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, *Non-local meets global: An integrated paradigm for hyperspectral image restoration*, IEEE Trans. Pattern Anal. Mach. Intell., 44 (2022), pp. 2089–2107, https://doi.org/10.1109/tpami.2020.3027563.

[17] X. He, J. Wu, Q. Ling, Z. Li, Z. Lin, and S. Zhou, *Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning*, IEEE Trans. Geosci. Remote Sens., 61 (2023), 5509917, https://doi.org/10.1109/tgrs.2023.3270667.

[18] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, *More diverse means better: Multimodal deep learning meets remote-sensing imagery classification*, IEEE Trans. Geosci. Remote Sens., 59 (2021), pp. 4340–4354, https://doi.org/10.1109/tgrs.2020.3016820.

[19] Z. Huang, L. Fang, and S. Li, *Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 58 (2020), pp. 5998–6007, https://doi.org/10.1109/tgrs.2019.2961703.

[20] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, *Hyperspectral anomaly detection via background and potential anomaly dictionaries construction*, IEEE Trans. Geosci. Remote Sens., 57 (2019), pp. 2263–2276, https://doi.org/10.1109/tgrs.2018.2872590.

[21] T. Jiang, Y. Li, W. Xie, and Q. Du, *Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 58 (2020), pp. 4666–4679, https://doi.org/10.1109/tgrs.2020.2965961.

[22] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, *Hyperspectral anomaly detection with attribute and edge-preserving filters*, IEEE Trans. Geosci. Remote Sens., 55 (2017), pp. 5600–5611, https://doi.org/10.1109/tgrs.2017.2710145.

[23] J. Kerekes, *Receiver operating characteristic curve confidence intervals and regions*, IEEE Geosci. Remote Sens. Lett., 5 (2008), pp. 251–255, https://doi.org/10.1109/lgrs.2008.915928.

[24] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, *Anomaly detection in hyperspectral images based on an adaptive support vector method*, IEEE Geosci. Remote Sens. Lett., 8 (2011), pp. 646–650, https://doi.org/10.1109/lgrs.2010.2098842.

[25] M. E. Kilmer and C. D. Martin, *Factorization strategies for third-order tensors*, Linear Algebra Appl., 435 (2011), pp. 641–658, https://doi.org/10.1016/j.laa.2010.09.020.

[26] D. Krishnan and R. Fergus, *Fast image deconvolution using hyper-Laplacian priors*, in Proceedings of the 22nd International Conference on Neural Information Processing Systems, NIPS'09, Curran Associates, Red Hook, NY, USA, 2009, pp. 1033–1041.

[27] H. Kwon and N. Nasrabadi, *Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery*, IEEE Trans. Geosci. Remote Sens., 43 (2005), pp. 388–397, https://doi.org/10.1109/tgrs.2004.841487.

[28] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, *Prior-based tensor approximation for anomaly detection in hyperspectral imagery*, IEEE Trans. Neural Netw. Learn. Syst., 33 (2022), pp. 1037–1050, https://doi.org/10.1109/tnnls.2020.3038659.

[29] W. Li, G. Wu, and Q. Du, *Transferred deep learning for anomaly detection in hyperspectral imagery*, IEEE Geosci. Remote Sens. Lett., 14 (2017), pp. 597–601, https://doi.org/10.1109/lgrs.2017.2657818.

[30] F. Liu and Q. Wang, *A sparse tensor-based classification method of hyperspectral image*, Signal Process., 168 (2020), 107361, https://doi.org/10.1016/j.sigpro.2019.107361.

[31] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, *Tensor robust principal component analysis with a new tensor nuclear norm*, IEEE Trans. Pattern Anal. Mach. Intell., 42 (2020), pp. 925–938, https://doi.org/10.1109/tpami.2019.2891760.

[32] G. Marjanovic and V. Solo, *On ℓ_q optimization and matrix completion*, IEEE Trans. Signal Proces., 60 (2012), pp. 5714–5724, https://doi.org/10.1109/tsp.2012.2212015.
[33] S. Matteoli, M. Diani, and G. Corsini, *Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images*, Opt. Eng., 49 (2010), 046201, https://doi.org/10.1117/1.3386069.
[34] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data*, Ann. Statist., 37 (2009), pp. 246–270, https://doi.org/10.1214/07-aos582.
[35] E. Merkurjev, A. Bertozzi, X. Yan, and K. Lerman, *Modified Cheeger and ratio cut methods using the Ginzburg-Landau functional for classification of high-dimensional data*, Inverse Problems, 33 (2017), 074003, https://doi.org/10.1088/1361-6420/33/7/074003.
[36] L. Pan and X. Chen, *Group sparse optimization for images recovery using capped folded concave functions*, SIAM J. Imaging Sci., 14 (2021), pp. 1–25, https://doi.org/10.1137/19m1304799.
[37] J. Peng, Y. Wang, H. Zhang, J. Wang, and D. Meng, *Exact decomposition of joint low rankness and local smoothness plus sparse matrices*, IEEE Trans. Pattern Anal. Mach. Intell., 45 (2023), pp. 5766–5781, https://doi.org/10.1109/tpami.2022.3204203.
[38] D. Qiu, M. Bai, M. K. Ng, and X. Zhang, *Nonlocal robust tensor recovery with nonconvex regularization*, Inverse Problems, 37 (2021), 035001, https://doi.org/10.1088/1361-6420/abd85b.
[39] D. Qiu, M. Bai, M. K. Ng, and X. Zhang, *Robust low-rank tensor completion via transformed tensor nuclear norm with total variation regularization*, Neurocomputing, 435 (2021), pp. 197–215, https://doi.org/10.1016/j.neucom.2020.12.110.
[40] I. Reed and X. Yu, *Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution*, IEEE Trans. Acoustics Speech Signal Process., 38 (1990), pp. 1760–1770, https://doi.org/10.1109/29.60107.
[41] S. Y. Rotfel'd, *The singular numbers of the sum of completely continuous operators*, in Spectral Theory, Vol. 3, Springer, Boston, MA, 1969, pp. 73–78, https://doi.org/10.1007/978-1-4684-7589-0_5.
[42] A. P.Schaum, *Hyperspectral anomaly detection beyond RX*, in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII, Vol. 6565, S. S. Shen and P. E. Lewis, eds., International Society for Optics and Photonics, Bellingham, WA, 2007, 656502, https://doi.org/10.1117/12.718789.
[43] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, *Tensor-based formulation and nuclear norm regularization for multienergy computed tomography*, IEEE Trans. Image Process., 23 (2014), pp. 1678–1693, https://doi.org/10.1109/tip.2014.2305840.
[44] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, *Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints*, IEEE Trans. Neural Netw. Learn. Syst., (2022), pp. 8679–8692, https://doi.org/10.1109/tnnls.2022.3152252.
[45] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, *Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery*, J. Appl. Remote Sens., 8 (2014), 083641, https://doi.org/10.1117/1.jrs.8.083641.
[46] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 58 (1996), pp. 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[47] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, *PDBSNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 61 (2023), 5511914, https://doi.org/10.1109/tgrs.2023.3276175.
[48] M. Wang, D. Hong, Y. Wu, L. Gao, and M. Huang, *Hyperspectral anomaly detection using tensor low-rank representation*, in Proceedings of 2022 10th China Conference on Command and Control, Springer, Singapore, 2022, pp. 127–132, https://doi.org/10.1007/978-981-19-6052-9_12.
[49] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, *Learning tensor low-rank representation for hyperspectral anomaly detection*, IEEE Trans. Cybernet., 53 (2023), pp. 679–691, https://doi.org/10.1109/tcyb.2022.3175771.
[50] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, *ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features*, IEEE Trans. Geosci. Remote Sens., 57 (2019), pp. 5146–5158, https://doi.org/10.1109/tgrs.2019.2897139.
[51] Y. Xu, *Alternating proximal gradient method for sparse nonnegative Tucker decomposition*, Math. Program. Comput., 7 (2014), pp. 39–70, https://doi.org/10.1007/s12532-014-0074-y.

[52] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, *Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA*, IEEE Trans. Geosci. Remote Sens., 56 (2018), pp. 2919–2930, https://doi.org/10.1109/tgrs.2017.2786718.

[53] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, *Anomaly detection in hyperspectral images based on low-rank and sparse representation*, IEEE Trans. Geosci. Remote Sens., 54 (2016), pp. 1990–2000, https://doi.org/10.1109/tgrs.2015.2493201.

[54] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imaging Sci., 6 (2013), pp. 1758–1789, https://doi.org/10.1137/120887795.

[55] Q. Yu and M. Yang, *Low-rank tensor recovery via non-convex regularization, structured factorization and spatio-temporal characteristics*, Pattern Recognit., 137 (2023), 109343, https://doi.org/10.1016/j.patcog.2023.109343.

[56] Q. Yu and X. Zhang, *A smoothing proximal gradient algorithm for matrix rank minimization problem*, Comput. Optim. Appl., 81 (2022), pp. 519–538, https://doi.org/10.1007/s10589-021-00337-9.

[57] Q. Yu, X. Zhang, Y. Chen, and L. Qi, *Low Tucker rank tensor completion using a symmetric block coordinate descent method*, Numer. Linear Algebra Appl., 30 (2023), e2464, https://doi.org/10.1002/nla.2464.

[58] Y. Yuan, D. Ma, and Q. Wang, *Hyperspectral anomaly detection by graph pixel selection*, IEEE Trans. Cybernet., 46 (2016), pp. 3123–3134, https://doi.org/10.1109/tcyb.2015.2497711.

[59] Y. Yuan, Q. Wang, and G. Zhu, *Fast hyperspectral anomaly detection via high-order 2-D crossing filter*, IEEE Trans. Geosci. Remote Sens., 53 (2015), pp. 620–630, https://doi.org/10.1109/tgrs.2014.2326654.

[60] C.-H. Zhang, *Nearly unbiased variable selection under minimax concave penalty*, Ann. Statist., 38 (2010), pp. 894–942, https://doi.org/10.1214/09-aos729.

[61] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, *Robust low-rank tensor recovery with rectification and alignment*, IEEE Trans. Pattern Anal. Mach. Intell., 43 (2021), pp. 238–255, https://doi.org/10.1109/tpami.2019.2929043.

[62] Y. Zhang, B. Du, L. Zhang, and S. Wang, *A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 54 (2016), pp. 1376–1389, https://doi.org/10.1109/tgrs.2015.2479299.

[63] C. Zhao, X. Li, and H. Zhu, *Hyperspectral anomaly detection based on stacked denoising autoencoders*, J. Appl. Remote Sens., 11 (2017), 042605, https://doi.org/10.1117/1.jrs.11.042605.

[64] X. Zhao, M. Bai, D. Sun, and L. Zheng, *Robust tensor completion: Equivalent surrogates, error bounds, and algorithms*, SIAM J. Imaging Sci., 15 (2022), pp. 625–669, https://doi.org/10.1137/21m1429539.

[65] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, *Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations*, IEEE Trans. Geosci. Remote Sens., 60 (2022), 5500117, https://doi.org/10.1109/tgrs.2020.3040221.