CONSTRAINED LOW-TUBAL-RANK TENSOR RECOVERY FOR HYPERSPECTRAL IMAGES MIXED NOISE REMOVAL BY BILATERAL RANDOM PROJECTIONS

Hao Zhang†, Xi-Le Zhao†∗, Tai-Xiang Jiang†, Michael Kwok-Po Ng‡

†School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, P. R. China
‡Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

ABSTRACT

In this paper, we propose a novel low-tubal-rank tensor recovery model, which directly constrains the tubal rank prior for effectively removing the mixed Gaussian and sparse noise in hyperspectral images. The constraints of tubal-rank and sparsity can govern the solution of the denoised tensor in the recovery procedure. To solve the constrained low-tubal-rank model, we develop an iterative algorithm based on bilateral random projections to efficiently solve the proposed model. The advantage of random projections is that the approximation of the low-tubal-rank tensor can be obtained quite accurately in an inexpensive manner. Experimental examples for hyperspectral image denoising are presented to demonstrate the effectiveness and efficiency of the proposed method.

Index Terms— Tensor, bilateral random projections, low-tubal-rank, mixed noise, hyperspectral images.

1. INTRODUCTION

Hyperspectral remote sensing images are widely used in various applications [1]. However, hyperspectral images (HSIs) in practice are often inevitably corrupted by several types of noises, such as Gaussian noise, sparse noise, stripes, and deadlines. Consequently, the applications are severely influenced. Therefore, it is essential to develop effective and efficient methods for HSIs denoising task like [2, 3].

HSIs are spatially and spectrally correlated resulting in low-rankness. Many denoising methods are devoted to preserving the low-rank structure of the clean HSIs. One classical way is unfolding HSIs to matrices, such as low-rank matrix recovery (LRMR) [4], nonconvex regularizer with weighted Schatten p-norm (WSNLRMA) [5], and low-rank subspace representation methods [6, 7]. However, unfolding HSIs to matrices will destroy the intrinsic structures.

Since tensor can better express more complex intrinsic structures of the higher-order data, the related researches have received considerable attention. A corrupted HSI is a three-way tensor, which can be decomposed as a clean part, a sparse

Fig. 1: Characteristics of the results recovered by CLTRTR. (a) Illustration of the number of non-zero tubes of the clean part, when the tubal rank prior is 3. (b) Illustration of the number of non-zero entries of the sparse noise part.

type part, and a Gaussian noise part (See Fig.1). The first and second dimensions of HSIs are corresponding to the spatial information while the third dimension reflects the spectral information. Spatial-spectral information can be simultaneously exploited by tensor-based methods. Based on different decomposition schemes, several low-rank tensor recovery methods have been proposed, such as low-$\tau$-rank tensor approximation (LRTA) [8] and the rank-1 tensor decomposition method [9].

The tensor tubal rank, based on the tensor singular value decomposition (t-SVD), is magnetic for well characterizing the inherent low-rank structure of a tensor. Tensor nuclear norm (TNN) was considered formulating a low-rank tensor recovery model (LRTR) [10] for effective HSIs denoising. Although TNN as a popular surrogate of the tubal rank has obtained promising results, it is just a suboptimal surrogate of the tubal rank. For the large size HSIs, LRTR suffers from the heavy burden of computing singular value decomposition (SVD) within each frontal slice. These motivate us to derive a novel constrained low-tubal-rank tensor recovery model with a fast and accurate algorithm for efficient HSIs mixed noise removal.

The main contributions of this work are two folds: (1) we propose a novel constrained low-tubal-rank tensor recovery model (CLTRTR) for effective HSI mixed Gaussian and
sparse noise removal, which directly constrains the tubal rank of the target HSI (Fig. 1 illustrates characteristics of the results generated by CLTRTR)); (2) we design a tensor bilateral random projections algorithm (t-BRP) to efficiently solve the proposed model, which finely and quickly approximates the low-tubal-rank tensor.

The rest of this paper is given as follows. Section 2 introduces notations and preliminaries. Section 3 gives the proposed model and algorithm. Section 4 reports the results of numerical experiments. Section 5 concludes this paper.

2. NOTATIONS AND PRELIMINARIES

For a three-way tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the $i$-th frontal slice is represented by $A^{(i)}$. We denote $\bar{\mathbf{A}}$ as the tensor generated by Discrete Fourier Transformation (DFT) on each tube of $\mathbf{A}$. We denote $\hat{\mathbf{A}}$ as a block diagonal matrix whose $i$-th block on the diagonal as the $i$-th frontal slice $A^{(i)}$ of $\bar{\mathbf{A}}$. Meanwhile, the block circulant matrix $\text{bcirc}(\mathbf{A})$ is defined as

$$
\text{bcirc}(\mathbf{A}) = \begin{pmatrix}
A^{(1)} & A^{(2)} & \cdots & A^{(n_3)} \\
A^{(2)} & A^{(1)} & \cdots & A^{(n_3-1)} \\
\vdots & \vdots & \ddots & \vdots \\
A^{(n_3)} & A^{(n_3-1)} & \cdots & A^{(1)}
\end{pmatrix}.
$$

The block circulant matrix can be block diagonalized, i.e.,

$$
(F_{n_3} \otimes I_{n_1}) \cdot \text{bcirc}(\mathbf{A}) \cdot (F_{n_3}^{-1} \otimes I_{n_2}) = \hat{\mathbf{A}},
$$

where $\otimes$ is the Kronecker product and $F_{n_3}$ is the DFT matrix.

Definition 2.1. (Tensor product) For $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{B} \in \mathbb{R}^{n_2 \times n_3}$, the tensor product is defined to be a $n_1 \times l \times n_3$ tensor $\mathbf{A} \ast \mathbf{B} = \text{fold} (\text{bcirc}(\mathbf{A}) \cdot \text{unfold}(\mathbf{B}))$. (2)

Definition 2.2. (Inverse of tensor) The inverse of a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is written as $\mathbf{A}^{-1}$, satisfying $\mathbf{A}^{-1} \ast \mathbf{A} = \mathbf{A} \ast \mathbf{A}^{-1} = \mathbf{I}$. $\mathbf{I}$ is the identity tensor whose first frontal slice is the $n \times n$ identity matrix.

Definition 2.3. (t-SVD) For $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor SVD of $\mathbf{A}$ can be factored as

$$
\mathbf{A} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^*,
$$

where $\mathbf{U} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times n_3}$ are orthogonal, $\mathbf{V}^*$ is the conjugate transpose of $\mathbf{V}$, and $\mathbf{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a $f$-diagonal tensor, whose each frontal slice is a diagonal matrix.

Definition 2.3. (Tensor multi-rank and tubal rank) The tensor multi-rank of $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $r \in \mathbb{R}^{n_3}$, whose $i$-th element is the rank of the $i$-th frontal slice of $\mathbf{A}$, i.e., $r_i = \text{rank}(A^{(i)})$. The tensor tubal rank, denoted as $\text{rank}_t(\mathbf{A})$, is defined as the number of non-zero tubes of $\mathbf{S}$, where $\mathbf{S}$ is from the t-SVD of $\mathbf{A} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^*$. That is

$$
\text{rank}_t(\mathbf{A}) = \# \{i : \mathbf{S}(i,i,:) \neq 0\} = \max r_i.
$$

Definition 2.4. (TNN) The TNN of a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathbf{A}\|_{\text{TNN}}$, is defined as the sum of singular values of all the frontal slices of $\bar{\mathbf{A}}$, i.e.,

$$
\|\mathbf{A}\|_{\text{TNN}} = \sum_{i=1}^{n_3} \|\bar{A}^{(i)}\|_*.
$$

It can be seen that the tubal rank is equal to the largest element of the multi-rank and TNN approximates the $l_1$-norm of the tensor multi-rank. Although TNN simplifies the recovery problem, it causes unavoidable bias.

Remark 2.1. For $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\text{rank}_t(\mathbf{A}) \leq \min(n_1, n_2)$ and $\text{rank}_t(\mathbf{A} \ast \mathbf{B}) \leq \min(\text{rank}_t(\mathbf{A}), \text{rank}_t(\mathbf{B}))$.

3. MAIN RESULTS

3.1. The proposed model

Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes an observed HSI, which can be expressed as the sum of three parts, i.e.,

$$
\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N},
$$

where $\mathbf{L}$ is the clean HSI, $\mathbf{S}$ is the sparse noise, and $\mathbf{N}$ is the Gaussian noise. The goal of HSI denoising is to recover $\mathbf{L}$ from observed $\mathbf{X}$. Assuming that $\mathbf{L}$ is low-tubal-rank with $\text{rank}_t(\mathbf{L}) \leq r$ and the number of non-zero elements of $\mathbf{S}$ is no more than $k$, i.e., $\text{card}(\mathbf{S}) \leq k$, we formulate the model

$$
\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2,
$$

s.t. $\text{rank}_t(\mathbf{L}) \leq r, \text{card}(\mathbf{S}) \leq k$. (7)

3.2. The proposed algorithm

Given $r$ and $k$, (7) can be transformed into solving the following two subproblems alternately until convergence:

$$
\mathbf{L}^t = \arg \min_{\text{rank}_t(\mathbf{L}) \leq r} \|\mathbf{X} - \mathbf{L} - \mathbf{S}^{t-1}\|_F^2,
$$

$$
\mathbf{S}^t = \arg \min_{\text{card}(\mathbf{S}) \leq k} \|\mathbf{X} - \mathbf{L}^t - \mathbf{S}\|_F^2.
$$

$\mathbf{L}^t$ can be obtained by t-SVD of $\mathbf{X} - \mathbf{S}^{t-1}$, due to the property that truncated t-SVD is optimal for data approximation. However, SVDs within each slice are the main computation burden at each iteration. To efficiently solve (8), we design a t-BRP algorithm, which approximates the truncated t-SVD.

Definition 3.1. (t-BRP) For $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-BRP of $\mathbf{X}$ can be constructed, i.e., $\mathbf{Y}_1 = \mathbf{X} \ast \mathbf{A}_1$, and $\mathbf{Y}_2 = \mathbf{X}^* \ast \mathbf{A}_2$, whereas $\mathbf{A}_1 \in \mathbb{R}^{n_2 \times n_3}$ and $\mathbf{A}_2 \in \mathbb{R}^{n_1 \times n_3}$ are random tensors.

The tubal rank-$r$ approximation of $\mathbf{X}$ can be conducted as

$$
\mathbf{L} = \mathbf{Y}_1 \ast (\mathbf{A}_1^* \ast \mathbf{Y}_1)^{-1} \ast \mathbf{Y}_2,
$$

where $\mathbf{A}_2^*$ is the conjugate transpose of $\mathbf{A}_2$ and (10) is an approximation of the truncated t-SVD as the explanation in (12).

Recalling (11) and (2), we construct the matrix bilateral random projections of each $\bar{X}^{(i)}$ as follows. For $i = 1, \ldots, n_3$, let $\bar{Y}_1^{(i)} = \mathbf{X}^{(i)} \bar{A}_1^{(i)}$, $\bar{Y}_2^{(i)} = (\mathbf{X}^{(i)} \ast \mathbf{A}_2^*)^{(i)}$, in which $\bar{A}_1^{(i)} \in \mathbb{C}^{n_2 \times r}$ and $\bar{A}_2^{(i)} \in \mathbb{C}^{n_1 \times r}$ are random matrices. The low-rank approximation of $\bar{X}^{(i)}$ is obtained by
\[ L^{(t)} = Y_1^{(t)}[\bar{A}_2^{(t)}]^*Y_2^{(t)} - 1 Y_2^{(t)}. \]
\[ \bar{Y}_1 = \bar{X}A_1, \bar{Y}_2 = \bar{X}^*\bar{A}_2, \text{ and} \]
\[ \bar{L} = \bar{Y}_1(\bar{A}_2^*\bar{Y}_2)^{-1}\bar{Y}_2. \] (11)

Then we have that \( (F_{n_3}^{-1} \otimes I_{n_2})X(F_{n_3}^{-1} \otimes I_{n_2}), (F_{n_3}^{-1} \otimes I_{n_2})A_1(F_{n_3}^{-1} \otimes I_{n_2}), (F_{n_3}^{-1} \otimes I_{n_2})X^*(F_{n_3}^{-1} \otimes I_{n_1}), (F_{n_3}^{-1} \otimes I_{n_2})A_2(F_{n_3}^{-1} \otimes I_{n_2}), [I_{n_1}^{-1}(F_{n_3}^{-1} \otimes I_{n_2})Y_1(F_{n_3}^{-1} \otimes I_{n_1})]^{-1}, \) and \( (F_{n_3}^{-1} \otimes I_{n_2})\bar{Y}_2 [F_{n_3}^{-1} \otimes I_{n_2}] \) are real block circulant matrices. We obtain expressions for \( bcirc(Y_1), bcirc(Y_2), \) and \( bcirc(L) \).

Folding up these results, the \( L \) can be given.

Considering a three-way tensor \( X \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) with rank\(_k(\mathcal{L}) \leq r \), the computational complexity of the developed t-BRP is \( O(n_1n_2n_3 \log n_3 + r^2(n_2 + r)n_3 + n_1n_2n_3r) \), while t-SVD is \( O(n_1n_2n_3 \log n_3 + \min(n_1n_2^2, n_1^2n_2)n_3) \). Referring to the Remark 2.1, the parameter \( r \) imposes a direct tubal rank constraint upon the estimated low-rank tensor \( \mathcal{L} \).

As for the subproblem \( \mathcal{S} \), \( \mathcal{S}^t \) is updated as
\[ S^t = \mathcal{H}_k(\mathcal{X} - \mathcal{L}^t), \] (12)
wherein \( \mathcal{H}_k(\mathcal{X} - \mathcal{L}^t) \) denotes the entry-wise hard thresholding operator, which sets all but the largest \( k \) elements of \( \mathcal{X} - \mathcal{L}^t \) to zero.

The proposed algorithm is summarized in Algorithm 1. The objective value \( \| \mathcal{X} - \mathcal{L} - \mathcal{S} \|_F^2 \) converges to a local minimum based on the framework in [13].

**Algorithm 1: CLTRTR for HSIs denoising.**

Input: \( \mathcal{X}, r, k, \epsilon \)
Output: \( \mathcal{L}, \mathcal{S} \)
Initialize: \( \mathcal{L}^0 = \mathcal{S}^0 = 0, t = 0 \)
While \( \| \mathcal{X} - \mathcal{L}^t - \mathcal{S}^t \|_F^2 / \| \mathcal{X} \|_F^2 > \epsilon \) do
\[ t = t+1; \]
\[ Y_1 = (\mathcal{X} - \mathcal{S}^{t-1})^*A_1, A_2 = Y_1, \]
\[ Y_2 = (\mathcal{X} - \mathcal{S}^{t-1})^{**}A_2; \]
If rank\(_k(A_2^*Y_2) \leq r \), then
\[ r = \text{rank}(A_2^*Y_2), \]
else continue
\[ \mathcal{L}^t = Y_1 + (A_2^*Y_1)^{-1}Y_2^2; \]
\[ \mathcal{S}^t = \mathcal{H}_k(\mathcal{X} - \mathcal{L}^t); \]
End while.

## 4. EXPERIMENTS

To illustrate the effectiveness of the proposed method, experiments are conducted on the synthetic and the real data. The compared methods consist of LRTA [8], BM4D [14], LRMR [4], WSNLRMA [5], and LRTR [10]. The parameters of the compared methods are optimally assigned or selected as suggested in the reference papers.

### 4.1. Synthetic data

The HSIs of Washington DC Mall (256 × 256 × 191) and Pavia University (610×340×103) are tested in the simulated experiments. The clean datasets are normalized to [0, 1] band-wise. Two noisy datasets are generated as follows.

http://lesun.weebly.com/hyperspectral-data-set.html

**Fig. 2:** Denoising results for band 79 of Washington DC Mall.

**Fig. 3:** Denoising results for band 79 of Pavia University.

**Case 1:** zero-mean Gaussian noise with variance \( \sigma = 0.04 \) and the sparse noise (salt and pepper) which affects 20% pixels are added to all bands of Washington DC Mall.

**Case 2:** zero-mean Gaussian noise with variance \( \sigma = 0.02 \) and the sparse noise which affects 20% pixels are added to all bands of Pavia University. Then stripes and deadlines are simultaneously added to 10 arbitrarily selected bands from this dataset with the width from one line to three lines.

For the proposed method, the parameters \( r \) and \( k \) are set to be 5 and \( 2 \times 10^6 \), respectively. The peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), and the spectral angle mapping (SAM) are employed for performance evaluation. Table I lists the quantitative comparisons. As observed, the proposed method outperforms the compared ones with respect to PSNR, SSIM, and SAM values. For the running time (in seconds), the proposed method is the third fastest, but considering the methods with results PSNR higher than 30dB, i.e., LRMR, WSNLRMA, and LRTR, the proposed method is the fastest.
The imagery of HYDICE urban data.

4.2. Real data

The imagery of HYDICE urban data is used in the experiment. The size of original dataset is $304 \times 304 \times 210$. The bands 104-108, 139-151, and 208-210 are removed due to serious pollution by the atmosphere and water absorption. The denoising results for band 103 are illustrated in Fig. 4. It can be seen that our method removes almost all stripes and finely preserves the intrinsic structure. LRTR effectively removes stripes, and WSNLRMA achieve great performance but destroy the partial-spatial structure. LRTRT effectively removes stripes, and time result contains evident loss of details.

5. CONCLUSION

In this paper, we introduced a novel constrained low-tubal-rank tensor recovery model to remove mixed noise in HSIs. Then we developed a t-BRP algorithm for efficiently solving the proposed model. The synthetic and real data experiments demonstrate that our method achieved excellent performance on HSIs mixed noise removal and preserved the structure. Moreover, the running time comparison indicated the high efficiency of our algorithm.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, and J. Chanussot, “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[2] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, “Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network,” *IEEE Trans. Geosci. Remote Sens.*, pp. 1–14, DOI:10.1109/TGRS.2018.2856197, 2018.

[3] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, “HSI-DeNet: Hyperspectral image restoration via convolutional neural network,” *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, DOI:10.1109/TGRS.2018.2859203, 2018.

[4] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, “Hyperspectral image restoration using low-rank matrix recovery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, 2014.

[5] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, “Hyperspectral image restoration via iteratively regularized weighted Schatten $p$-norm minimization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, 2016.

[6] L. Zhuang and J. M. Bioucas-Dias, “Fast hyperspectral image denoising based on low rank and sparse representations,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, 2018.

[7] T. Jiang, L. Zhuang, T. Huang, and J. M. Bioucas-Dias, “Adaptive hyperspectral mixed noise removal,” in *IGARSS*, pp. 4035–4038.

[8] N. Renard, S. Bourennane, and J. Blanc-Talon, “Denoising and dimensionality reduction using multilinear tools for hyperspectral images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, 2008.

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

[10] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, “Hyperspectral image restoration using low-rank tensor recovery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, 2017.

[11] Z. Zhang and S. Aeron, “Exact tensor completion using t-svd,” *IEEE Trans. Geosci. Remote Sens.*, vol. 65, no. 6, pp. 1511–1526, 2018.

[12] M. Fazel, E. Candes, B. Recht, and P. Parrilo, “Compressed sensing and robust recovery of low rank matrices,” in *ACSCC*, pp. 1043–1047.

[13] T. Zhou and D. Tao, “Godec: Randomized low-rank and sparse matrix decomposition in noisy case,” in *ICML*, pp. 33–40.

[14] M. Magni, V. Katkovnik, K. Egiazarian, and A. Foi, “Nonlocal transform-domain filter for volumetric data denoising and reconstruction,” *IEEE Trans. Image Processing.*, vol. 22, no. 1, pp. 119–133, 2013.

[15] J. M. Bioucas-Dias, A. Plaza, and J. Chanussot, “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[16] http://www.tec.army.mil/hypercube