

Towards Multi-Mode Outlier Robust Tensor Ring Decomposition

Yuning Qiu^{1,2}, Guoxu Zhou^{1,3}*, Andong Wang², Zhenhao Huang¹, Qibin Zhao^{2,1}

¹ School of Automation, Guangdong University of Technology, Guangzhou, 510006, China

² RIKEN Center for Advanced Intelligence Project, Tokyo, 1030027, Japan

³ Key Laboratory of Intelligent Detection and The Internet of Things in Manufacturing, Ministry of Education, Guangzhou, 510006, China

{yuning.qiu, andong.wang, qibin.zhao}@riken.jp, {gx.zhou}@gdut.edu.cn, zhhuang.gdut@qq.com

Abstract

Conventional Outlier Robust Tensor Decomposition (ORTD) approaches generally represent sparse outlier corruption within a specific mode. However, such an assumption, which may hold for matrices, proves inadequate when applied to high-order tensors. In the tensor domain, the outliers are prone to be corrupted in multiple modes simultaneously. Addressing this limitation, this study proposes a novel ORTD approach by recovering low-rank tensors contaminated by outliers spanning multiple modes. In particular, we conceptualize outliers within high-order tensors as latent tensor group sparsity by decomposing the corrupted tensor into a sum of multiple latent components, where each latent component is exclusive to outliers within a particular direction. Thus, it can effectively mitigate the outlier corruptions prevalent in high-order tensors across multiple modes. To theoretically guarantee recovery performance, we rigorously analyze a non-asymptotic upper bound of the estimation error for the proposed ORTD approach. In the optimization process, we develop an efficient alternate direction method of multipliers (ADMM) algorithm. Empirical validation of the approach’s efficacy is undertaken through comprehensive experimentation.

1 Introduction

Recent advancements in acquisition and processing for tensors (also referred to as multiway data) have facilitated their broad application across various fields. However, the disturbance on tensor data may amplify as the number of tensor orders increases. This is particularly evident in remote sensing imaging, where the data, influenced by multiple factors, often contains both ubiquitous low-amplitude noise and diverse outliers (Xu et al. 2015; Sun et al. 2019; Dao et al. 2021). Specifically, row-wise outliers stem from sensor scanning inconsistencies or hardware glitches, while column-wise outliers arise from abnormal sensor columns due to equipment issues or external interferences. Tube-wise outliers, spanning noise across spectral bands for a pixel, result from factors like sensor instability, hardware effects, or electromagnetic disruptions. These distinct outliers, whether appearing alone or together, present substantial hurdles to the integrity and interpretation of tensor data, underscoring

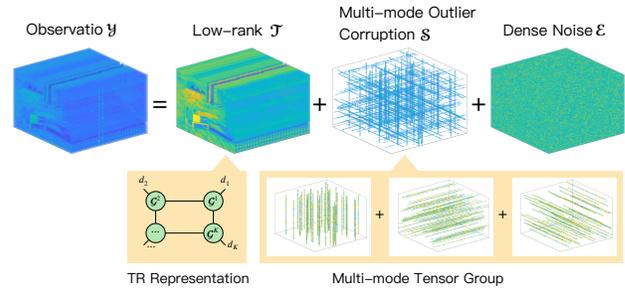


Figure 1: The problem addressed in this study: recover the “clean” low-rank tensor from multi-mode outlier and dense noise corrupted tensor.

the necessity for robust outlier resistance against multi-mode outliers.

The robust tensor decomposition (RTD) (Gu, Gui, and Han 2014; Anandkumar et al. 2016; Wang et al. 2020) provides a viable solution to noise removal and outlier resilience effectively (Dong et al. 2018; Cao et al. 2016; Xie et al. 2017). By modeling the corruption as a linear mixture of small dense noise and sparse outlier corruptions, the essence of RTD lies in distinguishing the inherent redundancy captured in the low-rank tensor from the significantly corrupted sparse outliers and dense noise present in high-order observations. Recent works have been extensively exploring the low-rank tensor within several tensor decompositions like CANDECOMP/PARAFAC decomposition (Carroll and Chang 1970; Harshman et al. 1970), Tucker decomposition (Tucker 1966), tensor train (TT) decomposition (Oseledets 2011), tensor ring (TR) decomposition (Zhao et al. 2016a), and tensor singular value decomposition (Kilmer and Martin 2011). Among the various tensor decompositions, tensor ring (TR) decomposition preserves circular shifting invariance, thereby capturing the intermodal redundancy present in real-world tensors and achieving success in numerous applications (Li and So 2021; Hou et al. 2019; Qiu et al. 2020; Peddireddy et al. 2022; Qiu et al. 2022a).

Nonetheless, contemporary advancements in RTD have primarily focused on strategies for capturing the low-rank structure within high-order tensor data, with relatively lim-

*Corresponding Author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ited emphasis on sparsity representation. This discrepancy is largely attributed to the stability of element-wise sparsity metrics in high-order tensors. In contrast, commonly utilized group sparsity techniques for outlier removal and detection lack the same favorable characteristics, as they can solely be defined for a single mode. Consequently, these approaches struggle to adequately outliers that span multiple modes within a tensor. To summarize, RTD proves effective in recovering low-rank tensors from data corrupted by outliers. However, these approaches for addressing outlier corruptions in multi-mode scenarios pose a burgeoning challenge. This prompts the following crucial questions:

Q1: Is it possible to formulate an Outlier Robust Tensor Decomposition (ORTD) approach capable of recovering low-rank tensors from multi-mode outlier corruptions?

Q2: Is there theoretical support for the estimation error of low-rank and sparse outlier tensors?

To answer these questions, our study introduces a novel approach called multi-mode outlier-robust tensor ring decomposition (ORTRD) for the recovery of low-rank tensors from noisy observations. The key innovation lies in our utilization of multi-mode tensor group sparsity (as defined in Definition 4) to present outliers in high-order tensors, allowing for their presence across multiple directions. In conjunction with the classical tensor decompositions, we employ a TR structure to capture the underlying low-rank structure. To ensure reliable recovery performance, we rigorously analyze the deterministic upper bound (Theorem 1) and non-asymptotic upper bound (Theorem 2) of the estimation error. Our comprehensive investigation demonstrates that the proposed approach achieves significant improvements over conventional robust matrix/tensor decomposition techniques. This enhanced capability makes ORTRD exceptionally versatile in representing low-rank tensors that are subjected to outlier corruptions across multiple modes (refer to Figure 1). In summary, our work contributes in the following ways:

- To estimate the clean low-rank tensors, we develop a novel ORTD approach by incorporating a newly established multi-mode tensor group sparsity alongside a TR rank representation.
- We theoretically reveal a non-asymptotic upper bound for the estimation error to justify the potential effectiveness of the proposed ORTRD approach.
- The proposed ORTRD is versatile for various tensor data recovery tasks, including synthetic low-rank tensor data, hyperspectral videos (HSVs), and light field images. Extensive experiments validate the superiority of the proposed approach compared to state-of-the-art approaches.

2 Notations and Preliminaries

First, we introduce necessary notations in this paper. We set the scalars to be the lowercase letters, e.g., $a \in \mathbb{R}$, vectors to be the bold lowercase letter, e.g., $\mathbf{a} \in \mathbb{R}^d$, and matrices be the bold upper-case letters, e.g., $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$, and tensors be the Euler script letters, e.g., $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. We use c_1, c_2 , and so forth, to represent universal constants whose values might change from one context to another. The (i_1, i_2, \dots, i_K) th

entry of tensor \mathcal{X} is given by $\mathcal{X}(i_1, i_2, \dots, i_K)$, and i th column of matrix \mathbf{X} is denoted by $\mathbf{X}(:, i)$. We let the set of all positive integers $1, 2, \dots, K$ be $[K]$. We let the nuclear norm of a matrix \mathbf{X} be $\|\mathbf{X}\|_*$. The $\ell_{2,1}$ norm and its dual norm of a matrix \mathbf{X} is defined by $\|\mathbf{X}\|_{2,1}$ and $\|\mathbf{X}\|_{2,\infty}$, respectively. The mode- k unfolding (or matricization) (Kolda and Bader 2009) of a tensor \mathcal{X} is given by $\mathbf{X}_{(k)}$.

Next, we delve into several pivotal concepts of TR decomposition.

Definition 1 (Tensor Ring (TR) Decomposition (Zhao et al. 2016a)). *Given an arbitrary K th-order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$, TR decomposition represents it by the circular contraction over K third-order tensors as $\mathcal{X} = \text{TR}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \dots, \mathcal{G}^{(K)})$, where core tensors $\mathcal{G}^{(k)} \in \mathbb{R}^{r_k \times d_k \times r_{k+1}}$, $k \in [K]$, and $r_{K+1} = r_1$. Alternatively, it can be represented as the following element-wise formulation:*

$$\mathcal{X}(i_1, i_2, \dots, i_K) = \sum_{u_1, u_2, \dots, u_K}^{r_1, r_2, \dots, r_K} \prod_{k=1}^K \mathcal{G}^{(k)}(u_k, i_k, u_{k+1}), \quad (1)$$

where the vector $[r_1, r_2, \dots, r_K]$ is denoted as TR rank.

Low-rank TR finds applications in diverse fields such as signal/image processing (Li and So 2021), quantum physics (Peddireddy et al. 2022), machine learning (Hou et al. 2019), and more, enabling efficient analysis and manipulation of high-order data. For handling low-rank TR structures, the circular unfolding of tensors provides an effective approach to leverage matrix analysis techniques.

Definition 2 (Tensor Circular Unfolding (Yu et al. 2018, 2019)). *Given an arbitrary K th-order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$, its mode- k circular unfolding matrix is $\mathbf{X}_{\langle k \rangle} \in \mathbb{R}^{d_{1,k} \times d_{2,k}}$, where $d_{1,k} = \prod_{u=l+1}^{k-1} d_u$ and $d_{2,k} = \prod_{u=k}^l d_u$, and $s = \lceil K/2 \rceil$ denotes the number of indexes maintain in $d_{2,k}$, and l is given by*

$$l = \begin{cases} k + s - 1, & k + s \leq K, \\ k + s - 1 - K, & \text{otherwise.} \end{cases} \quad (2)$$

Alternatively, its element-wise form is denoted by $\mathcal{X}(i_1, i_2, \dots, i_K) = \mathbf{X}_{\langle k \rangle}(i_{l+1}, \dots, i_{k-1}, i_k, \dots, i_l)$.

Lemma 1. (Yu et al. 2019) *Given a K th-order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$ with TR rank $[r_1, r_2, \dots, r_K]$, the rank of circular unfolding matrix $\mathbf{X}_{\langle k \rangle}$ is bounded by $r_k r_{k+s}$.*

Lemma 1 allows us to relax the intractable TR rank minimization problem to K matrix rank minimization problems. By utilizing its convex surrogate, the matrix nuclear norm, to replace the rank, we proceed to introduce the Tensor Ring Nuclear Norm (TRNN), a powerful tool for effectively utilizing the low-rank TR structures in various applications (Yu et al. 2019; Huang et al. 2020a; He and Atia 2022; Zhang, Zhang, and Wang 2022).

Definition 3. (Yu et al. 2019) *Given a K th-order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$, its TRNN is defined by $\|\mathcal{X}\|_{\text{trnn}} := \sum_{k=1}^K \alpha_k \|\mathbf{X}_{\langle k \rangle}\|_*$, where $\alpha_k > 0$, $k \in [K]$ and $\sum_{k=1}^K \alpha_k = 1$ denotes the weight parameter.*

Lemma 2. (Qiu et al. 2022b) Given an arbitrary K th-order tensor \mathcal{X} , the dual norm of its TRNN is given by $\|\mathcal{X}\|_{\text{trnn}}^* = \inf_{\mathbf{x}=\mathbf{x}^1+\mathbf{x}^2+\dots+\mathbf{x}^K} \max_{k \in [K]} \alpha_k^{-1} \|\mathbf{X}_{\langle k \rangle}^k\|$, where $\|\cdot\|$ denotes the matrix spectral norm, and \mathcal{X}^k is the k th latent tensor component.

Lemma 2 plays a crucial role in the analysis of the statistical performance of the proposed approach.

3 Multi-mode Outlier Robust Tensor Ring Decomposition

Suppose that there is an observed corrupted tensor $\mathcal{Y} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$. We assume that the intrinsic ‘‘clean’’ low-rank tensor has a TR representation, and the sparse tensor contains outlier corruptions across multiple modes. Thus, the observation problem can be given by

$$\mathcal{Y} = \mathcal{T} + \mathcal{S} + \mathcal{E}, \quad (3)$$

where \mathcal{T} and \mathcal{S} represent the underlying low-rank tensor and the outlier tensor, respectively, and \mathcal{E} is dense noise tensor. We provide insight into the underlying low-rank tensor \mathcal{T} , which is relatively weak and quite reasonable for many real-world recorded data like images and videos:

Assumption 1. The ℓ_∞ norm of the low-rank tensor \mathcal{T} is not spiky, i.e., $\varphi := \|\mathcal{T}\|_\infty < +\infty$.

Subsequently, we formally present the challenge of multi-mode outlier-robust tensor decomposition: *How can we proficiently reconstruct an unknown tensor \mathcal{T} from the observed data \mathcal{Y} that is tainted by multi-mode outliers \mathcal{S} and dense noises \mathcal{E} ?*

To address the challenge, it’s crucial to consider the multi-mode nature of the unknown outliers. In light of this multi-mode characteristic, we propose a novel norm to measure the complexity of the outlier tensor \mathcal{S} in Eq. (3). The objective is to represent the unknown multi-mode outliers by a specific linear mixture of all single-mode sparse outliers that retains the minimal overall group sparsity.

Definition 4 (Multi-mode Tensor Group Sparsity, MTGS). Given an arbitrary K th-order tensor \mathcal{S} , the multi-mode tensor group sparsity can be given by the latent tensor decomposition approach, where each latent component contains outliers from one mode:

$$\|\mathcal{S}\|_{\text{mtgs}} := \inf_{\mathcal{S}=\mathcal{S}^1+\mathcal{S}^2+\dots+\mathcal{S}^K} \sum_{k=1}^K \beta_k \|\mathcal{S}_{(k)}^k\|_{2,1}, \quad (4)$$

where the infimum is taken over the K tuple of latent tensors $\mathcal{S}^1, \dots, \mathcal{S}^K$ that sums to \mathcal{S} , and $\mathcal{S}_{(k)}^k$ is the mode- k unfolding of the k th latent component \mathcal{S}^k , $\beta_k > 0$ denotes the weight parameter.

To be consistent with our key question Q1, we further assume the outlier tensor \mathcal{S} may contain multi-mode outliers and further satisfy the following incoherence assumption¹.

¹The multi-mode assumption makes our observation model (3) significantly different from existing works like (Gu, Gui, and Han 2014; Anandkumar et al. 2016; Wang et al. 2020)

Assumption 2. The $\ell_{2,\infty}$ norm of the l th unfolding of each latent component \mathcal{S}^k of the outlier tensor \mathcal{S} is bounded as $\|\mathcal{S}_{(l)}^k\|_{2,\infty} \leq \varrho$, for all $l, k \in [K]$, $l \neq k$.

Consequently, under TRNN and MTGS regularizations on low-rank tensor and multi-mode outliers, we have the following ORTRD approach:

$$\begin{aligned} \min_{\mathcal{T}, \mathcal{S}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{T} - \mathcal{S}\|_{\text{F}}^2 + \lambda \|\mathcal{T}\|_{\text{trnn}} + \tau \|\mathcal{S}\|_{\text{mtgs}}, \\ \text{s.t.} \quad & \|\mathcal{T}\|_\infty \leq \varphi, \|\mathcal{S}_{(l)}^k\|_{2,\infty} \leq \varrho, l \neq k, \end{aligned} \quad (5)$$

where $\lambda > 0$ and $\tau > 0$ are penalty parameters. The optimization algorithm employed to solve the Eq. (5) hinges on the utilization of the Alternating Direction Method of Multipliers (ADMM) algorithm (Boyd et al. 2011). For a comprehensive understanding, please refer to Appendix B.

4 Theoretical Recovery Performance

To address the key question Q2, we theoretically investigate the statistical performance of the ORTRD approach.

Let $\{\hat{\mathcal{T}}, \hat{\mathcal{S}}, \hat{\mathcal{S}}^{k,*}\}$ and $\{\mathcal{T}^*, \mathcal{S}^*, \mathcal{S}^{k,*}\}$ be the estimated tensor and intrinsic true tensors, respectively. The residual tensor is given by $\Delta\mathcal{T} := \mathcal{T}^* - \hat{\mathcal{T}}$, $\Delta\mathcal{S} := \mathcal{S}^* - \hat{\mathcal{S}}$ and $\Delta\mathcal{S}^{k,*} := \mathcal{S}^{k,*} - \hat{\mathcal{S}}^{k,*}$. We let $\Omega^k \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$ be the support set of outliers in the k th latent component \mathcal{S}^k , $\Omega = \sum_{k=1}^K \Omega^k$ be the support set of \mathcal{S} , and $|\Omega_{(k)}^k|$ be the number of corrupted columns. To clearly show the recovery performance of the proposed approach, we first present the deterministic bound in the following theorem.

Theorem 1. If we set the penalty parameters by $\lambda \geq \|\mathcal{E}\|_{\text{trnn}}^*$ and $\tau \geq \max_k \|\mathbf{E}_{(k)}\|_{2,\infty} / \beta_k + \max_k 2D\varphi / (\sqrt{d_k}\beta_k) + \max_k (K-1)\varrho / \beta_k$, then any optimal solution of the proposed ORTRD satisfies

$$\begin{aligned} \|\Delta\mathcal{T}\|_{\text{F}}^2 + \sum_{k=1}^K \|\Delta\mathcal{S}^k\|_{\text{F}}^2 &\leq c_1 \lambda^2 \sum_{k=1}^K \alpha_k^2 r_k r_{k+s} + c_2 \tau^2 |\hat{\Omega}|, \\ \|\Delta\mathcal{T}\|_{\text{F}}^2 + \|\Delta\mathcal{S}\|_{\text{F}}^2 &\leq c_3 \lambda^2 \sum_{k=1}^K \alpha_k^2 r_k r_{k+s} + c_4 \tau^2 \beta_{k'}^2 |\Omega_{(k')}|, \end{aligned}$$

where $|\hat{\Omega}| = \sum_{k=1}^K \beta_k^2 |\Omega_{(k)}^k|$ and $k' = \arg \min_k \|\mathcal{S}_{(k)}\|_{2,1}$.

Proof of Theorem 1 is given in Appendix A.1. Theorem 1 indicates that when penalty parameters (λ, τ) exceed some thresholds, then the estimation error of low-rank tensor \mathcal{T} and structure sparse tensor \mathcal{S} is upper bound by the sum of latent structure sparsity $|\Omega^k|$ and the sum of the square of TR rank r^2 . However, the upper bound of $\ell_{2,\infty}$ norm of $\mathbf{E}_{(k)}$ and spectral norm of \mathcal{E} is still unknown. To clearly show its upper bound, we assume that the noise tensor $\mathcal{E} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$ follows *i.i.d.* Gaussian distribution $\mathcal{N}(0, \sigma^2)$, and then present a non-asymptotic upper bound on the estimation error in the following theorem.

Theorem 2. For Gaussian noise tensor \mathcal{E} , if the penalty parameters $\lambda = 2\sigma K^{-2} \sum_{k=1}^K (\sqrt{d_{1,k}} + \sqrt{d_{2,k}}) / \alpha_k$ and $\tau = \max_k \sigma (\sqrt{d_k} + 3\sqrt{\log D}) / \beta_k + \max_k 2D\varphi / (\sqrt{d_k}\beta_k) +$

$\max_k (K-1)\varrho/\beta_k$, then with high probability, any optimal solution of the proposed ORTRD satisfies:

$$\begin{aligned} & \|\Delta\mathcal{J}\|_F^2 + \sum_{k=1}^K \|\Delta\mathcal{S}^k\|_F^2 \leq \\ & \frac{c_5\sigma^2}{K^4} \sum_{k=1}^K \left(\frac{\tilde{d}_{1,2}^k}{\alpha_k}\right)^2 \sum_{k=1}^K \alpha_k^2 r_k r_{k+s} + c_6\tau^2 \sum_{k=1}^K \beta_k^2 |\Omega_{(k)}^k|, \quad (6) \\ & \|\Delta\mathcal{J}\|_F^2 + \|\Delta\mathcal{S}\|_F^2 \leq \\ & \frac{c_7\sigma^2}{K^4} \sum_{k=1}^K \left(\frac{\tilde{d}_{1,2}^k}{\alpha_k}\right)^2 \sum_{k=1}^K \alpha_k^2 r_k r_{k+s} + c_8\tau^2 \beta_{k'}^2 |\Omega_{(k')}|, \end{aligned}$$

where $\tilde{d}_{1,2}^k = \sqrt{d_{1,k}} + \sqrt{d_{2,k}}, \forall k \in [K]$.

The proof of Theorem 2 is presented in Appendix A.2. Compared with the exact subspace recovery guarantee in ORPCA or OTRPCA (Xu, Caramanis, and Sanghavi 2012; Zhou and Feng 2017), the established theorem relaxes the strong matrix/tensor incoherence condition on the low-rank matrix/tensor. Moreover, Theorem 2 gives the estimation error bound on the recovered low-rank and outlier corrupted tensor, while ORPCA or OTRPCA can only guarantee the subspace recovery on the low-rank matrix/tensor.

By letting $d_k = d$, $\alpha_k = 1/K$ and $\beta_k = \beta$, we can observe that the upper bound of the estimation error Eq. (6) in Theorem 2 is established in the following oracle inequality with high probability

$$\frac{\|\Delta\mathcal{J}\|_F^2}{D} + \frac{\|\Delta\mathcal{S}\|_F^2}{D} \lesssim \sigma^2 \left(\frac{r^2}{d^{\lfloor K/2 \rfloor}} + \xi^2 \frac{|\Omega_{(k')}|}{D} \right), \quad (7)$$

where \lesssim denotes that the inequality holds up to a multiplicative absolute constant and a factor, $\xi := \sqrt{d} + 3\sqrt{\log D}$. From Eq. (7), it is clear that Theorem 2 shows substantial improvements compared with the matrix's and tensor's case (Klopp, Lounici, and Tsybakov 2017; Zhou and Feng 2017; Wang, Jin, and Tang 2020). Theorem 2 breaks down the rank limitation of $r \leq d$ in classical low-rank recovery tasks. This is because, even if $r \geq d$, a sufficiently small upper bound for the estimation error can still be achieved due to the possibility that $r^2 \leq d^{\lfloor K/2 \rfloor}$. This contrasts with previous robust matrix/tensor decomposition approaches where low-rank recovery typically falters once $r \geq d$. Moreover, the estimation error of outlier components is determined by the minimum number of corrupted columns in k th-mode, i.e., $|\Omega_{(k')}|$. The result is equivalent to mode-specific ORTD approaches by simply setting the outlier direction to k' (Klopp, Lounici, and Tsybakov 2017; Wang, Jin, and Tang 2020). However, the proposed approach does not require this prior, which demonstrates the flexibility of the proposed approach.

5 Experimental Results

In this section, we evaluate the performance of the proposed approach by conducting experiments on both synthetic data and real-world datasets, including light field images and hyperspectral videos. We compare the experimental results with some state-of-the-art robust matrix/tensor decomposition approaches, including, RPCA (Candès

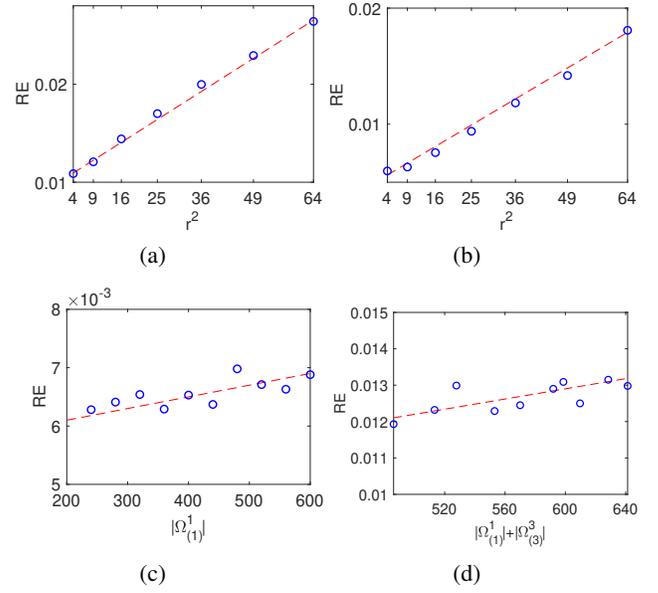


Figure 2: RE versus the number of outliers: (a) when outlier corruptions are distributed in the first mode; (b) when outlier corruptions are distributed in the first and the third modes. RE versus the square of TR rank: (c) when outlier corruptions are distributed in the first mode; (d) when outlier corruptions are distributed in the first and the third modes.

et al. 2011), ORPCA (Xu, Caramanis, and Sanghavi 2010), BRTF (Zhao et al. 2016b), HoRPCA (Goldfarb and Qin 2014), RTRC (Huang et al. 2020b), TTRPCA (Yang et al. 2020), and ORTPCA (Zhou and Feng 2017). We evaluate the performance of the compared approaches in terms of the relative squared error (RSE), which is defined by $\|\mathcal{J}^* - \hat{\mathcal{J}}\|_F^2 / \|\mathcal{J}^*\|_F^2$. The implementation code is available at <https://github.com/ynqiu/MORTRD>.

5.1 Verifying the Correctness of Theorem 2

In this part, we conduct synthetic low-rank tensor recovery experiments to investigate the correctness of the established upper bound in Theorem 2. To generate synthetic low rank tensor $\mathcal{J}^* \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$ with TR rank $[r_1, r_2, \dots, r_K]$, we first generated K core tensors $\mathcal{G}^{(k)} \in \mathbb{R}^{r_k \times d_k \times r_{k+1}}$ where each entry is produced by the *i.i.d.* Gaussian distribution $\mathcal{N}(0, 1)$. To construct the latent structural tensor $\mathcal{S}^{k,*}$, we let the support set of $\mathcal{S}^{k,*}$ be Ω^k , and then randomly select $|\Omega_{(k)}^k|$ columns of $\mathcal{S}^{k,*}$ as outliers whose entries obey *i.i.d.* $\mathcal{N}(0, 1)$. Thus, the outlier is given by $\mathcal{S} = \sum_{k=1}^K \mathcal{S}^{k,*}$. The additive noise tensor is produced by $\mathcal{N}(0, \sigma^2)$, where $\sigma = 10^{-3} \|\mathcal{J}^*\|_F / \sqrt{D}$ to guarantee a constant signal-to-noise ratio (SNR). All the experiments are repeated 10 times and their mean values are reported.

According to Eq. (7), the upper bound of the entry-wise estimation error of our proposed approach is linearly scaled with variables $r^2/d^{\lfloor K/2 \rfloor}$ and $|\Omega_{(k')}|/D$. Thus, by fixing one of these two variables, and linearly increasing the other one,

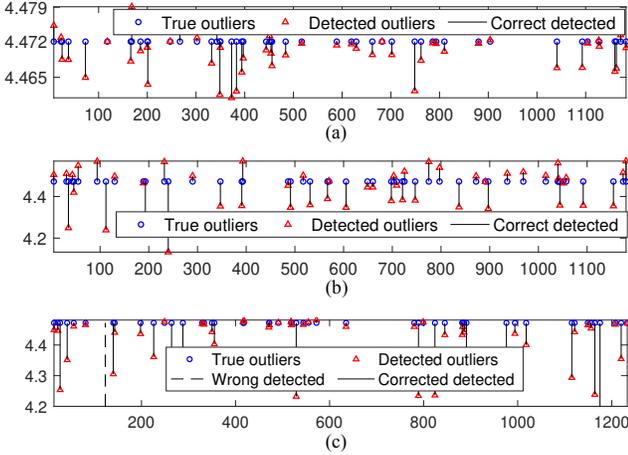


Figure 3: Column-wise Frobenius norm versus column index. (a) Frobenius norm versus column index when outliers are distributed across the first mode; (b) Frobenius norm versus column index on the first mode when outliers are distributed across the first and the third modes; (c) Frobenius norm versus column index on the third mode when outliers are distributed across the first and the third modes.

the estimation error should also increase linearly. We adopt reconstruction error $RE = (\|\Delta\mathcal{J}\|_F^2 + \|\Delta\mathcal{S}\|_F^2)/D$ as the evaluation metric. We first investigate RE value versus TR rank. We generate the synthetic tensor of size $20 \times 20 \times 20 \times 20$ with TR rank $r \in \{2, 3, \dots, 8\}$. In Figure 2 (a)-(b), we depict the RE versus the square of TR rank when outliers are distributed in the first mode, and in both the first and the third modes. It can be observed that the RE value increases linearly with r^2 , which verifies the results in Eq. (7). In Figure 2 (c)-(d), we plot the RE versus $|\Omega_{(1)}^1|$ and $|\Omega_{(1)}^1| + |\Omega_{(3)}^3|$, respectively. Similarly, the estimation error increases linearly with the minimum number of non-zero fibers in multi-mode outlier corrupted tensor, which is consistent with the theoretical results in Eqs (6) and (7).

5.2 Verifying the Effectiveness of Outlier Detection

The approach for generating low-rank and outlier corruptions is the same as the preceding section. The low-rank tensor is of size $20 \times 20 \times 20 \times 20$ with TR rank $r = 3$. The outlier corruptions are distributed on the first mode $\mathcal{S} = \mathcal{S}^1$ with $|\Omega_{(1)}^1| = 240$. We compute column-wise Frobenius norm of $\hat{\mathcal{S}}_{(1)}^1$ and $\mathcal{S}_{(1)}^{1,*}$, and visualize the first 50 outliers in Figure 3(a). We can easily find that all the outliers are successfully detected by the proposed approach. Subsequently, we conduct the experiment on multiple latent components by generating outlier tensors $\mathcal{S} = \mathcal{S}^1 + \mathcal{S}^3$ with $|\Omega_{(1)}^1| = |\Omega_{(3)}^3| = 240$. Figure 3(b)-(c) depict the column-wise Frobenius norm on $\mathcal{S}_{(1)}^1$ and $\mathcal{S}_{(3)}^3$, respectively. All of the outliers in the first mode are successfully detected, but there is one column in $\mathcal{S}_{(3)}^3$ being misclassified as outliers,

HSV data	Outlier modes	ORPCA	RTRC	ORTPCA	Ours
bus	1	0.0755	0.0464	0.0488	0.0452
	1,2	0.3712	0.0555	0.0600	0.0450
	1,2,3	0.5210	0.0578	0.0665	0.0451
	1,2,3,4	0.6385	0.0603	0.0713	0.0452
coin	1	0.1180	0.0449	0.0911	0.0417
	1,2	0.7354	0.0890	0.1184	0.0415
	1,2,3	1.0401	0.0908	0.1328	0.0416
	1,2,3,4	1.2892	0.0934	0.1471	0.0416
face	1	0.0773	0.0458	0.0438	0.0448
	1,2	0.4226	0.0548	0.0557	0.0445
	1,2,3	0.5953	0.0564	0.0638	0.0446
	1,2,3,4	0.7308	0.0583	0.0695	0.0445
hand	1	0.0710	0.0461	0.0407	0.0456
	1,2	0.3435	0.0529	0.0516	0.0454
	1,2,3	0.4823	0.0544	0.0587	0.0455
	1,2,3,4	0.5908	0.0559	0.0643	0.0453

Table 1: Comparison of RSE values of compared approaches on 4 HSVs. **Best** in boldface.

which demonstrates the effectiveness of the proposed approach in detecting multi-mode outlier corruptions.

5.3 Hyperspectral Video Recovery

In this part, we compare the recovery results for all the approaches. We randomly select four HSV datasets². These HSVs can be viewed as fourth-order tensors of size $64 \times d_2 \times 16 \times 50$, where d_2 is distinct from each video. The entries in each spectral are rescaled in an interval $[0, 1]$. The additive Gaussian noise is set as Section 5.1 with $\sigma = 0.05 \|\mathcal{J}^*\|_F / \sqrt{D}$. The multi-mode outliers are generated with $|\Omega_{(k)}^k| = \text{round}(10^{-2} \prod_{j,j \neq k} d_j)$, and each outlier $\mathcal{S}_{(k)}^k(:, i)$ is generated by a uniform discrete distribution on $[-1, 1]$.

Table 1 gives RSE values of different compared methods on four HSV datasets under varying outlier corruptions. Due to space limitations, only part of the results are depicted. The full comparisons can be obtained in Appendix C.1. The proposed approach obtains the lowest RSE in most cases, and RTRC obtains the second-best performance, which verifies the superiority of the proposed multi-mode robust tensor decomposition framework and low-rank approximation ability of TR representation. ORTPCA and ORPCA also obtain comparable performance when outliers only lie on the first mode, but the performance degenerates as the number of outlier modes increases. Thus, these mode-specific outlier robust matrix/tensor decomposition is effective when outliers exactly match the specific column space.

5.4 Light Field Image Recovery and Outlier Detection

In this part, we investigate recovery performance and outliers detection results of all compared methods. We adopt a publicly accessible light field images dataset³, and ran-

²<https://www.hsitracking.com/contest/>

³<https://lightfield-analysis.uni-konstanz.de/>

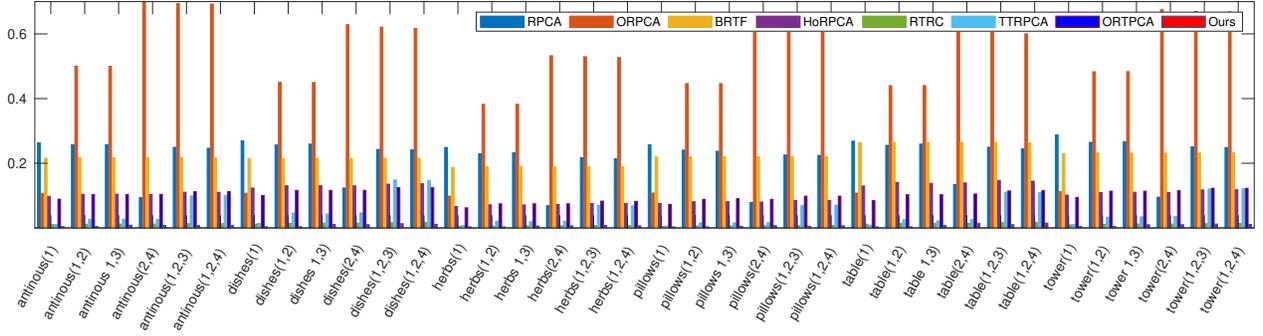


Figure 4: Comparison of the RSE values of 8 compared methods for image denoising on 6 light field images. The numbers behind the light field images denote the outliers aligned in different modes.

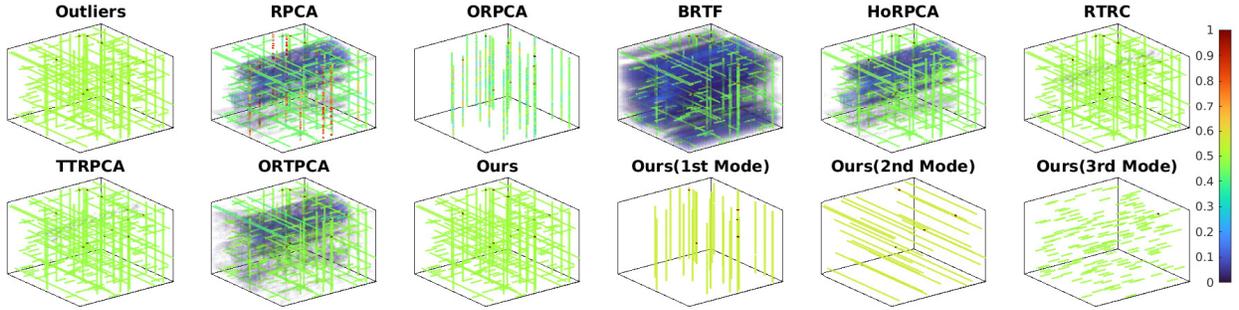


Figure 5: The visual comparison of the recovered sparse outlier tensors by different methods.

domly select 6 of these images. These images can be viewed as fourth-order tensors of size $64 \times 64 \times 9 \times 9$. The support set of outlier corruptions are generated with $|\Omega_{(k)}^k| = \text{round}(10^{-3} \prod_{j,j \neq k} d_j)$. The remaining settings are set as the HSV recovery section.

Figure 4 depicts the RSE values of the compared approaches on 6 light field images. From the comparisons, we can observe that: 1) The proposed method achieves the lowest RSE in all tested images and all experiment settings, which indicates the effectiveness of the proposed method; 2) Both RTRC and TTRPCA achieve comparable recovery performance, demonstrating the superiority of low-TT/TR-rank approximation for high-order tensor recovery; 3) As the number of outlier modes increases, the RSE of all the compared methods shows slight increases; 4) When outliers only appear in the first mode, ORPCA also obtains comparable recovery results. However, when outliers appear in multiple modes, ORPCA performs worse since it can only estimate outliers from one pre-fixed mode.

Subsequently, we investigate the outlier detection results. For the compared methods, we adopt Kmeans clustering to group all $\|\mathbf{S}_{(k)}(:, i)\|_F^2$ into two classes (outliers vs. non-outliers) for outlier detection. The proposed approach automatically separates outliers into different latent components, thus we can directly cluster $\|\mathbf{S}_{(k)}^k(:, i)\|_F^2$ into two classes. We repeat 10 times clustering to eliminate the local convergence effect of Kmeans. Then, we present the average clustering accuracy from all corrupted modes in Table 2. We

can find that the proposed method obtains the best clustering accuracy in all cases. Besides, when the outliers are only distributed in the spatial dimensions, i.e., the first and the second modes, most of the compared methods obtain 100% clustering accuracy. To elucidate the factors leading to the observed outlier detection results, we visually present the detected sparse outliers for an intuitive interpretation with $|\Omega_{(k)}^k| = \text{round}(10^{-3} \prod_{j,j \neq k} d_j), k \in \{1, 2, 3\}$. Then we reshape the outlier tensors into third-order tensors of size $64 \times 64 \times 81$ and then visualize them in Figure 5. It can be observed that the proposed method can accurately separate the outlier tensor whereas other methods are either incapable of recovering outliers across multiple modes (as is the case with ORPCA), or yield sparse components that significantly deviate from the original outliers (as is the case with RPCA, BRTF, HoRPCA, and ORTPCA). Consequently, these results verify that the proposed method is adept at effectively segregating outliers corrupted from multiple directions.

5.5 Convergence Behaviour

In this part, we investigate the convergence behavior of the proposed algorithm on light field images. We plot the relative change of the recovered low-rank tensor in Figure 6(a) on *antinous* and *bedroom* datasets, where we let $|\Omega_{(k)}^k| = (10^{-2} \prod_{j,j \neq k} d_j), k \in \{1, 2, 3\}$. It can be observed that the proposed algorithm usually converges within 50 iterations. In Figure 6(b)-(d), we plot the relative change of each latent component by setting the outliers modes $k = 1, k \in \{1, 2\}$

Data	Outlier modes	ORPCA	RTRC	ORTPCA	Ours
<i>antinous</i>	1	100	100	100	100
	1,3	80.64	82.04	82.02	99.70
	1,2,3	72.53	85.55	85.56	99.75
<i>dishes</i>	1	100	100	100	100
	1,3	81.28	82.62	82.59	99.61
	1,2,3	71.88	85.55	85.58	99.73
<i>herbs</i>	1	100	100	100	100
	1,3	81.28	82.62	82.62	99.65
	1,2,3	72.14	85.55	85.55	99.72
<i>pillows</i>	1	100	100	100	100
	1,3	80.87	82.17	82.15	99.74
	1,2,3	71.67	85.55	85.57	99.76
<i>table</i>	1	100	100	100	100
	1,3	81.54	82.62	82.56	99.57
	1,2,3	72.12	85.55	85.60	99.69
<i>tower</i>	1	100	100	100	100
	1,3	80.38	82.17	82.16	99.64
	1,2,3	71.56	85.55	85.56	99.70

Table 2: Outliers detection ACC (%) on compared methods

and $k \in \{1, 2, 3\}$, respectively. In cases where a given mode is devoid of outliers, the relative change of associated latent components shows a zig-zag from iterations 20 to 60 but finally converges the NaN. This is reasonable since the latent component will approach zeros if the corresponding mode does not contain any outlier corruptions. These results further substantiate the capability of our approach to automatically identify the corrupted modes from high-order tensors.

6 Related Works

While RTD is often seen as an extension of robust principal component analysis (RPCA) for higher-order tensors, making such extensions isn't straightforward due to the intricate multilinear structures inherent in high-order tensor data. Unlike matrices, definitions of tensor rank are not straightforward and come with variations, stemming primarily from tensor decomposition approaches introduced in the Introduction section. To characterize the sparse corruption for high-order tensor, the convex element-wise sparsity measure, i.e., ℓ_1 norm, stands out as one of the most extensively employed strategies (Goldfarb and Qin 2014; Lu et al. 2020), since it is far easier to be extended from vector or matrix to high-order tensor. However, it breaks down when outliers are distributed column-wise, i.e., large errors concentrate only on a number of fibers rather than scattering uniformly across the sparse component. One alternative solution is to unfold the high-order tensor into a matrix along the outliers so that they can be well captured with $\ell_{2,1}$ norm (Hu and Work 2020; Zhou and Feng 2017). These methods approximate the intrinsic low-rank tensor with a well-defined tensor decomposition approach and matrix-based group sparsity penalty (Xu, Caramanis, and Sanghavi 2010).

Some typical research works for the ORTD are developed recently. (Zhang et al. 2014) introduced a TRPCA method based on the tensor nuclear norm (TNN) and $\ell_{1,1,2}$ norm on a third-order tensor, where the $\|\mathcal{S}\|_{\ell_{1,1,2}} := \sum_{i,j} \|\mathcal{S}(i, j, :)\|_2$.

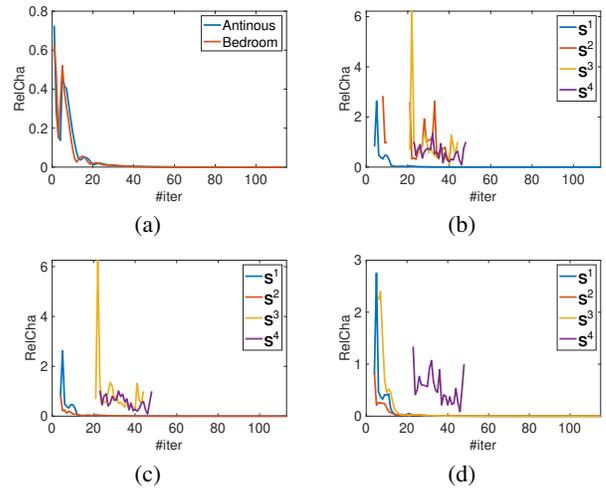


Figure 6: Convergence curves on light field images. (a) The relative change of the low-rank tensor \mathcal{J} on *antinous* and *bedroom* images when outlier corruptions distributed in $[1, 2, 3]$ modes; (b)-(d) The relative change of $\{\mathcal{S}^k\}_{k=1}^4$ on *Antinous* image when outlier corruptions distributed in $[1]$, $[1, 2]$ and $[1, 2, 3]$ modes, respectively.

In fact, $\ell_{1,1,2}$ norm is equivalent to the $\ell_{2,1}$ norm on mode-3 unfolding of a third-order tensor. To capture the slice-wise outliers, (Zhou and Feng 2017) proposed an outlier TRPCA approach by adopting TNN and $\ell_{2,1}$ norm on a third-order tensor, where $\|\mathcal{S}\|_{\ell_{2,1}} := \sum_j \mathcal{S}(:, j, :)$. These methods target outliers of higher-order tensors but encounter significant challenges: (i) Prior mode knowledge is necessary for outlier distribution, and (ii) the $\ell_{2,1}$ norm focuses on one mode, failing to capture multi-mode outliers effectively.

7 Conclusions and Discussions

This work proposes Outlier Robust Tensor Ring Decomposition (ORTRD) - a novel method to reliably decompose tensors corrupted by outliers. The key innovation is a new multi-mode tensor group sparsity norm that effectively models outlier corruption along multiple tensor modes. Leveraging this norm, we formulate a robust tensor low-rank framework decomposing the corrupted tensor into clean low-rank and sparse outlier components. We have proven ORTRD can recover underlying low-rank and sparse tensors with estimation error bounded under mild conditions. Comprehensive experiments on both synthetic and real-world data have validated our theoretical analysis and demonstrated that ORTRD can successfully recover clean low-rank structures from tensors corrupted by multi-mode outliers.

Limitations First, the proposed solver for TRNN minimization based on SVDs may face challenges in scaling to large-size data, and designing more scalable algorithms will be an important focus for future work. Second, as the proposed convex model may not provide tight enough relaxations to approximate non-convex rank and sparsity functions, investigating non-convex regularizers for tighter relaxations will be another valuable direction for future research.

Acknowledgments

Yuning Qiu was supported by the RIKEN Special Postdoctoral Researcher Program. This work was supported by the National Natural Science Foundation of China under Grant 62203124, 62073087, 62071132, and 62103110, and in part by the Guangdong Natural Science Foundation under Grant 2023A1515012916.

References

- Anandkumar, A.; Jain, P.; Shi, Y.; and Niranjan, U. N. 2016. Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations. In *Artificial Intelligence and Statistics*, 268–276. PMLR.
- Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1–122.
- Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3): 1–37.
- Cao, W.; Wang, Y.; Sun, J.; Meng, D.; Yang, C.; Cichocki, A.; and Xu, Z. 2016. Total variation regularized tensor RPCA for background subtraction from compressive measurements. *IEEE Transactions on Image Processing*, 25(9): 4075–4090.
- Carroll, J. D.; and Chang, J.-J. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3): 283–319.
- Dao, P. D.; Mantripragada, K.; He, Y.; and Qureshi, F. Z. 2021. Improving hyperspectral image segmentation by applying inverse noise weighting and outlier removal for optimal scale selection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171: 348–366.
- Dong, W.; Huang, T.; Shi, G.; Ma, Y.; and Li, X. 2018. Robust tensor approximation with Laplacian scale mixture modeling for multiframe image and video denoising. *IEEE Journal of Selected Topics in Signal Processing*, 12(6): 1435–1448.
- Goldfarb, D.; and Qin, Z. T. 2014. Robust Low-Rank Tensor Recovery: Models and Algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1): 225–253.
- Gu, Q.; Gui, H.; and Han, J. 2014. Robust tensor decomposition with gross corruption. *Advances in Neural Information Processing Systems*, 27.
- Harshman, R. A.; et al. 1970. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis.
- He, Y.; and Atia, G. K. 2022. Multi-Mode Tensor Space Clustering Based on Low-Tensor-Rank Representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 6893–6901.
- Hou, M.; Tang, J.; Zhang, J.; Kong, W.; and Zhao, Q. 2019. Deep multimodal multilinear fusion with high-order polynomial pooling. *Advances in Neural Information Processing Systems*, 32.
- Hu, Y.; and Work, D. B. 2020. Robust tensor recovery with fiber outliers for traffic events. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(1): 1–27.
- Huang, H.; Liu, Y.; Liu, J.; and Zhu, C. 2020a. Provable tensor ring completion. *Signal Processing*, 171: 107486.
- Huang, H.; Liu, Y.; Long, Z.; and Zhu, C. 2020b. Robust low-rank tensor ring completion. *IEEE Transactions on Computational Imaging*, 6: 1117–1126.
- Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.
- Klopp, O.; Lounici, K.; and Tsybakov, A. B. 2017. Robust matrix completion. *Probability Theory and Related Fields*, 169: 523–564.
- Kolda, T. G.; and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500.
- Li, X. P.; and So, H. C. 2021. Robust Low-Rank Tensor Completion Based on Tensor Ring Rank via $\ell_{p,\epsilon}$ -Norm. *IEEE Transactions on Signal Processing*, 69: 3685–3698.
- Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2020. Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4): 925–938.
- Oseledets, I. V. 2011. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5): 2295–2317.
- Peddireddy, D.; Bansal, V.; Jacob, Z.; and Aggarwal, V. 2022. Tensor ring parametrized variational quantum circuits for large scale quantum machine learning. *arXiv preprint arXiv:2201.08878*.
- Qiu, Y.; Zhou, G.; Wang, Y.; Zhang, Y.; and Xie, S. 2020. A generalized graph regularized non-negative Tucker decomposition framework for tensor data representation. *IEEE transactions on cybernetics*, 52(1): 594–607.
- Qiu, Y.; Zhou, G.; Zeng, J.; Zhao, Q.; and Xie, S. 2022a. Imbalanced low-rank tensor completion via latent matrix factorization. *Neural Networks*, 155: 369–382.
- Qiu, Y.; Zhou, G.; Zhao, Q.; and Xie, S. 2022b. Noisy Tensor Completion via Low-Rank Tensor Ring. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.
- Sun, W.; Yang, G.; Peng, J.; and Du, Q. 2019. Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(1): 107–111.
- Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3): 279–311.
- Wang, A.; Jin, Z.; and Tang, G. 2020. Robust tensor decomposition via t-SVD: Near-optimal statistical guarantee and scalable algorithms. *Signal Processing*, 167: 107319.
- Wang, A.; Li, C.; Jin, Z.; and Zhao, Q. 2020. Robust tensor decomposition via orientation invariant tubal nuclear norms. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04): 6102–6109.
- Xie, K.; Li, X.; Wang, X.; Xie, G.; Wen, J.; Cao, J.; and Zhang, D. 2017. Fast tensor factorization for accurate internet anomaly detection. *IEEE/ACM transactions on networking*, 25(6): 3794–3807.

- Xu, H.; Caramanis, C.; and Sanghavi, S. 2010. Robust PCA via outlier pursuit. *Advances in neural information processing systems*, 23.
- Xu, H.; Caramanis, C.; and Sanghavi, S. 2012. Robust PCA via Outlier Pursuit. *IEEE Transactions on Information Theory*, 58(5): 3047–3064.
- Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; and Wei, Z. 2015. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4): 1990–2000.
- Yang, J.-H.; Zhao, X.-L.; Ji, T.-Y.; Ma, T.-H.; and Huang, T.-Z. 2020. Low-rank tensor train for tensor robust principal component analysis. *Applied Mathematics and Computation*, 367: 124783.
- Yu, J.; Li, C.; Zhao, Q.; and Zhao, G. 2019. Tensoring nuclear norm minimization and application for visual: Data completion. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 3142–3146. IEEE.
- Yu, J.; Zhou, G.; Zhao, Q.; and Xie, K. 2018. An effective tensor completion method based on multi-linear tensor ring decomposition. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1344–1349. IEEE.
- Zhang, Y.; Zhang, Y.; and Wang, W. 2022. Learning Linear and Nonlinear Low-Rank Structure in Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering*.
- Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3842–3849.
- Zhao, Q.; Zhou, G.; Xie, S.; Zhang, L.; and Cichocki, A. 2016a. Tensor ring decomposition. *arXiv preprint arXiv:1606.05535*.
- Zhao, Q.; Zhou, G.; Zhang, L.; Cichocki, A.; and Amari, S.-I. 2016b. Bayesian Robust Tensor Factorization for Incomplete Multiway Data. *IEEE Transactions on Neural Networks and Learning Systems*, 27(4): 736–748.
- Zhou, P.; and Feng, J. 2017. Outlier-Robust Tensor PCA. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3938–3946. Honolulu, HI: IEEE. ISBN 978-1-5386-0457-1.