Refining Dual Spectral Sparsity in Transformed Tensor Singular Values

Anonymous Author(s)
Affiliation
Address
email

Abstract

The Tensor Nuclear Norm (TNN), derived from the tensor Singular Value Decomposition, is a central low-rank modeling tool that enforces *element-wise sparsity* on frequency-domain singular values and has been widely used in multi-way data recovery for machine learning and computer vision. However, as a direct extension of the matrix nuclear norm, it inherits the assumption of *single-level spectral sparsity*, which strictly limits its ability to capture the *multi-level spectral structures* inherent in real-world data—particularly the coexistence of low-rankness within and sparsity across frequency components. To address this, we propose the tensor ℓ_p-Schatten-q quasi-norm ($p, q \in (0, 1]$), a new metric that enables *dual spectral sparsity control* by jointly regularizing both types of structure. While this formulation generalizes TNN and unifies existing methods such as the tensor Schatten-p norm and tensor average rank, it differs fundamentally in modeling principle by coupling global frequency sparsity with local spectral low-rankness. This coupling introduces significant theoretical and algorithmic challenges. To tackle these challenges, we provide a theoretical characterization by establishing the first minimax error bounds under dual spectral sparsity, and an algorithmic solution by designing an efficient reweighted optimization scheme tailored to the resulting nonconvex structure. Numerical experiments demonstrate the effectiveness of our method in modeling complex multi-way data.

1 Introduction

Modeling latent structural patterns in high-dimensional signals is a fundamental challenge across domains such as machine learning and signal processing [17, 38, 19]. Real-world datasets are often inherently multi-modal and high-dimensional (tensor-form), containing intricate dependencies that cannot be adequately captured by naïve modeling or vector/matrix-based representations [4]. A common strategy to uncover these relationships is to impose a *low-rank* prior, which isolates essential information and reduces the degrees of freedom, focusing on the principal components of the signal [21, 1]. Traditional tensor decomposition methods, such as CANDECOMP/PARAFAC (CP) [3], Tucker [27], and Tensor Train [23], have been widely used to model tensor signals [4, 16, 8, 34]. While effective in certain scenarios, these methods rely on the assumption of intrinsic low-rankness in the *original domain*, which may fail to hold in complex, real-world applications. This limitation has led to the development of *transformed-domain* modeling, where linear transformations like the Discrete Fourier Transform (DFT) are applied to reveal more pronounced low-rank patterns. Within this paradigm, the tensor Singular Value Decomposition (t-SVD) has emerged as a powerful framework with notable success in applications such as image and video analysis [17, 38, 32, 30].

Building on the t-SVD framework, the Tensor Nuclear Norm (TNN) has become an extensively adopted regularizer for low-rank tensor modeling [20, 38, 25, 6, 36, 18, 39]. By extending the matrix

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

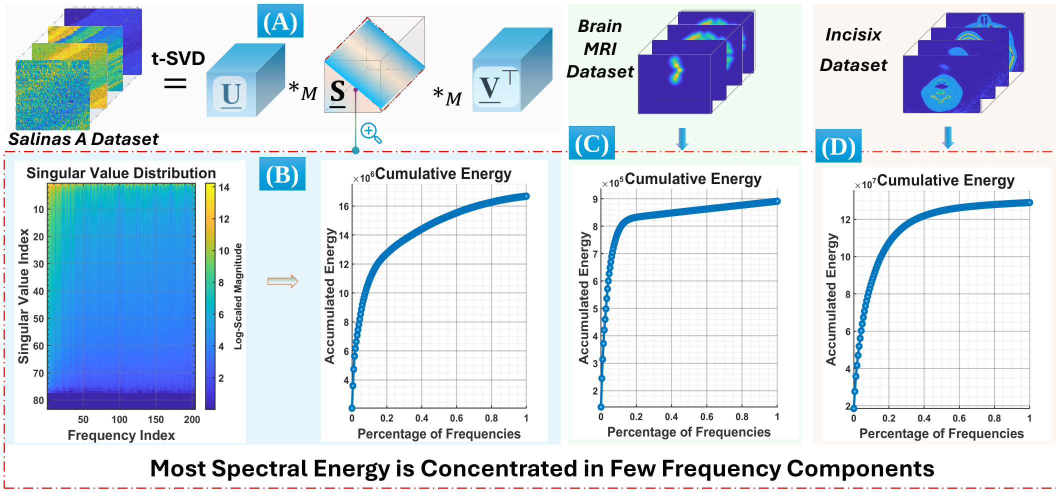

Figure 1: Empirical illustration of dual spectral sparsity patterns in the transformed (DCT) domain via t-SVD. (A) The t-SVD framework decomposes a tensor into frequency-domain singular structures. (B)-Left: Singular value heatmap of the *Salinas A* dataset under t-SVD—each column represents one frequency slice. Vertical decay reveals intra-frequency low-rankness, while horizontal variation indicates sparsity across frequencies. (B)-Right, (C), (D): Cumulative energy curves for *Salinas A*, *Brain MRI*, and *Incisix* datasets show that over 80% of total spectral energy is concentrated in the top 15%–30% frequency components, confirming frequency-wise sparsity. These observations support the presence of a dual-level spectral structure and motivate regularization schemes that go beyond uniform norms like TNN [38, 19] to jointly model frequency sparsity and low-rankness.

nuclear norm to the tensor setting, TNN promotes low-rankness by enforcing *element-wise sparsity* on singular values in the transformed domain [13, 38]. This formulation effectively captures low-rank dependencies within individual frequency components.

However, a long-overlooked limitation of TNN lies in its assumption of *uniform spectral regularization*, which treats all frequency components equally regardless of their relative importance. From a signal processing perspective, this *single-level sparsity* design fails to account for the dual-level structure often observed in transformed tensor data. In particular, real-world tensors may exhibit strong low-rankness within each frequency component along with sparsity across the frequency domain. As illustrated in Fig.1 and further discussed in §3, empirical analyses of several datasets, including hyperspectral images and medical imaging volumes, indicate that a small subset of frequency slices contributes the majority of spectral energy. In addition, these dominant components often exhibit pronounced low-rank structures within each frequency slice. These observations suggest the need for a more flexible regularization framework that can separately characterize both intra-frequency low-rankness and inter-frequency sparsity, instead of relying on a uniform scheme like TNN.

These limitations necessitate a new method capable of modeling both levels of sparsity. This raises three interconnected questions:

RQ1 (Modeling): *how to effectively model both intra-frequency and inter-frequency dependencies in tensor data?*

RQ2 (Theory): *can we establish rigorous theoretical guarantees to validate such a framework, given the challenges of analyzing coupled sparsity?*

RQ3 (Algorithm): *can efficient algorithms be designed to tackle the optimization challenges introduced by the coupled sparsity structure?*

To address these questions, we propose the *tensor ℓ_p-Schatten-q quasi-norm*, a novel framework introducing *dual spectral sparsity control* to simultaneously model both within-frequency and across-frequency dependencies. Specifically, parameter p governs sparsity among different frequency components (**RQ1**), while parameter q controls low-rankness within each frequency component. This framework generalizes and extends TNN, unifying existing methods such as the tensor Schatten-p quasi-norm [12] and tensor average rank [31] into a single, versatile framework.

While our framework offers promising modeling capabilities, the *coupled nature of this dual spectral sparsity* introduces significant theoretical and computational challenges. Our main contributions in developing and validating this framework are as follows:

- **Structural Modeling (RQ1):** To the best of our knowledge, this work is the first to rigorously formalize and explicitly model a coupled spectral structure within the t-SVD framework, where inter-frequency sparsity coexists with intra-frequency low-rankness (Section 3). The proposed ℓ_p-Schatten-q quasi-norm jointly models both inter-frequency sparsity and intra-frequency low-rankness, while allowing separate control over each via parameters p and q. This formulation captures hierarchical spectral structure beyond uniform regularizers such as TNN.
- **Theoretical Guarantees (RQ2):** We establish sharp minimax lower and upper bounds for tensor estimation under dual spectral sparsity, covering both hard and soft regimes (Section 4). The analysis introduces new techniques to characterize the complexity of coupled parameter spaces, extending classical tools such as covering numbers and metric entropy to the tensor spectral setting.
- **Optimization and Empirical Validation (RQ3):** We develop a scalable proximal algorithm tailored to the proposed quasi-norm (Section 5). It employs a reweighted $\ell_{1/2}$ approximation and frequency-wise singular value updates in the transform domain, effectively handling the nonconvexity and structural coupling induced by dual spectral sparsity. Experiments on real-world tensor recovery tasks demonstrate the potential applicability of our method (Section 6).

The remainder of the paper is organized as follows. Section 2 reviews basic preliminaries. Section 3 introduces the proposed quasi-norm. Sections 4 and 5 present the theoretical analysis and optimization algorithm, respectively. Experimental results are reported in Section 6, followed by the conclusion in Section 7. Details on related work, proofs, algorithms, and experiments are provided in the appendix.

2 Notations and Preliminaries

Notations. For any positive integer d, let $[d] := \{1, \ldots, d\}$. We denote vectors by lowercase bold letters (e.g., $\mathbf{a}$), matrices by uppercase bold letters (e.g., $\mathbf{A}$), and 3-way tensors by underlined uppercase letters (e.g., $\underline{\mathbf{A}}$). Constants, represented as c and its variants (e.g., c_1, C), may vary in value across contexts. For a 3-way tensor of size $d_1 \times d_2 \times m$, we assume $d_1 \geq d_2$ without loss of generality.

For a matrix $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$, we define $\boldsymbol{\sigma}(\mathbf{A})$ as the vector of its singular values, arranged in descending order. The spectral norm $\|\mathbf{A}\|_{\text{spec}}$ and nuclear norm $\|\mathbf{A}\|_*$ of $\mathbf{A}$ are defined as the largest and the sum of its singular values, respectively. For any tensor $\underline{\mathbf{A}}$, we define its ℓ_p-norm as $\|\underline{\mathbf{A}}\|_p := \|\operatorname{vec}(\underline{\mathbf{A}})\|_p$ and its Frobenius norm as $\|\underline{\mathbf{A}}\|_{\mathrm{F}} := \|\operatorname{vec}(\underline{\mathbf{A}})\|_2$, where $\operatorname{vec}(\cdot)$ denotes the vectorization operation [11]. The inner product of two tensors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ is given by $\langle \underline{\mathbf{A}}, \underline{\mathbf{B}} \rangle := \operatorname{vec}(\underline{\mathbf{A}})^\top \operatorname{vec}(\underline{\mathbf{B}})$. For a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{d_1 \times d_2 \times m}$, we denote its i-th frontal slice as $\underline{\mathbf{A}}_{:,:,i}$ or simply $\underline{\mathbf{A}}_i$ when clear from context.

The t-SVD Framework. The t-SVD framework is based on the t-product operation, a generalization of matrix multiplication to tensors, which operates under an invertible linear transform M [9]. By enhancing low-rank properties through specific linear transformations, this approach effectively exploits intrinsic correlations within the data [36, 29]. This paper adopts the convention of using orthogonal matrices for M due to their stability and computational advantages [18, 28]. Specifically, for an orthogonal matrix $\mathbf{M} \in \mathbb{R}^{m \times m}$, we define the M-linear transform and its inverse on a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ as:

$$M(\underline{\mathbf{T}}) := \underline{\mathbf{T}} \times_3 \mathbf{M}, \quad \text{and} \quad M^{-1}(\underline{\mathbf{T}}) := \underline{\mathbf{T}} \times_3 \mathbf{M}^{-1}, \tag{1}$$

where $\times_3$ denotes the mode-3 tensor-matrix product [9]. Using this transform, we introduce the basic notions in the t-SVD framework.

Definition 2.1 (t-product [9])**.** The t-product of two tensors $\underline{\mathbf{A}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{d_2 \times d_3 \times m}$ under the transform M in (1) is denoted by $\underline{\mathbf{A}} *_M \underline{\mathbf{B}} = \underline{\mathbf{C}} \in \mathbb{R}^{d_1 \times d_3 \times m}$, where $M(\underline{\mathbf{C}}) = M(\underline{\mathbf{A}}) \odot M(\underline{\mathbf{B}})$ in the transformed domain, and $\odot$ denotes the frontal-slice-wise product of the tensors.

Definition 2.2 (M-block-diagonal matrix [28])**.** For a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$, its M-block-diagonal matrix $\bar{\mathbf{T}} \in \mathbb{R}^{d_1 m \times d_2 m}$ is defined as

$$\bar{\mathbf{T}} := \operatorname{bdiag}(M(\underline{\mathbf{T}})) = \operatorname{diag}\left(M(\underline{\mathbf{T}})_{:,:,1}, \ldots, M(\underline{\mathbf{T}})_{:,:,m}\right),$$

where $M(\underline{\mathbf{T}})$ is the mode-3 transform of $\underline{\mathbf{T}}$, and the operator $\operatorname{bdiag}(\cdot)$ stacks the frontal slices as diagonal blocks.

We now formally introduce the t-SVD, as illustrated in Fig. 1-(A).

Definition 2.3 (t-SVD and tensor tubal rank [9])**.** The tensor Singular Value Decomposition (t-SVD) of a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ under the invertible linear transform M in (1) is:

$$\underline{\mathbf{T}} = \underline{\mathbf{U}} *_M \underline{\mathbf{S}} *_M \underline{\mathbf{V}}^\top, \tag{2}$$

where $\underline{\mathbf{U}} \in \mathbb{R}^{d_1 \times d_1 \times m}$ and $\underline{\mathbf{V}} \in \mathbb{R}^{d_2 \times d_2 \times m}$ are t-orthogonal tensors, and $\underline{\mathbf{S}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ is an f-diagonal tensor. The tubal rank of $\underline{\mathbf{T}}$ is defined as the number of non-zero tubes in $\underline{\mathbf{S}}$ in the t-SVD, i.e., $r_{\text{tb}}(\underline{\mathbf{T}}) := \#\{i \mid \underline{\mathbf{S}}_{i,i,:} \neq \mathbf{0}, i \leq \min\{d_1, d_2\}\}$.

To further model the low-rank structure of tensors in the transformed domain, the tensor nuclear norm (TNN) is proposed as a key regularizer in low-rank tensor learning:

Definition 2.4 (Tensor nuclear norm [20])**.** The tensor nuclear norm (TNN) of a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ under the transform M are defined as $\|\underline{\mathbf{T}}\|_* := \|\bar{\mathbf{T}}\|_* = \|\boldsymbol{\sigma}(\bar{\mathbf{T}})\|_1$.

In this definition, TNN captures *the element-wise sparsity of the transformed spectrum* $\boldsymbol{\sigma}(\bar{\mathbf{T}}) \in \mathbb{R}^{m \cdot \min\{d_1, d_2\}}$, allowing it to promote low-rank characteristics in the spectral domain. This property has made TNN a foundational tool in tensor analysis, particularly for low-rank tensor recovery in various applications such as image inpainting [19].

3 Dual Spectral Sparsity in the t-SVD Framework

Effectively capturing both intra-frequency low-rankness and inter-frequency sparsity (**RQ1**) is essential for modeling structured tensor data. While methods like TNN emphasize within-frequency low-rankness, they overlook sparsity across frequencies, limiting their ability to represent hierarchical dependencies. To overcome this, we introduce the ℓ_p-Schatten-q quasi-norm, a dual-sparsity regularization framework designed to capture both levels of structure in a unified way.

Limitations of TNN from a Group Sparsity Perspective. According to Definition 2.4, the tensor nuclear norm (TNN) promotes low-rankness by enforcing element-wise sparsity on singular values in the transformed domain, effectively capturing intra-frequency low-rank structures. However, it applies uniform regularization across all frequency components, regardless of their spectral importance. This design fails to exploit the potential sparsity across frequency slices that is often present in real-world tensors. Fig. 1 presents empirical evidence from three representative datasets—*Salinas A*[1], *Brain MRI* [33], and *Incisix* [5]—demonstrating that only a small portion of frequency components accounts for the majority of spectral energy. Specifically, more than 80% of the energy is concentrated in the top 15%–30% of frequency bands. Meanwhile, the singular value heatmap (Fig. 1(B)-Left) reveals pronounced horizontal sparsity, indicating that many frequency slices contribute minimally. Within each active frequency slice, singular values decay rapidly, confirming low-rankness.

These observations suggest a dual-level structure comprising inter-frequency sparsity and intra-frequency low-rankness. From a group sparsity perspective, the spectrum $\boldsymbol{\sigma}(\bar{\mathbf{T}})$ can be partitioned into groups, where each group corresponds to the singular values $\boldsymbol{\sigma}(M(\underline{\mathbf{T}})_{:,:,i})$ of a specific frequency slice. TNN enforces uniform regularization across these groups, overlooking their heterogeneous importance. As a result, it may underperform when modeling data with hierarchical spectral structures. These limitations motivate a more expressive framework that separately accounts for both levels of structure.

Hard Dual Spectral Sparsity. To address the limitations of TNN, we first define a hard dual spectral sparsity structure, where the tensor is assumed to satisfy exact sparsity constraints across and within frequency components. This serves as an idealized formulation that captures the extreme case of dual spectral sparsity and provides a clean theoretical foundation for later analysis.

Definition 3.1 (Hard Dual Spectral Sparsity)**.** A tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$ is said to exhibit (s, r)-dual sparsity under a linear transform M if it satisfies two constraints:

I. *Inter-frequency sparsity:* The number of active frequency components is limited to at most s. Specifically, only s out of the m frequency components can have non-zero singular value vectors: $\sum_{i=1}^m \mathbb{I}\left(\boldsymbol{\sigma}(M(\underline{\mathbf{T}})_{:,:,i}) \neq \mathbf{0}\right) \leq s$, where $\boldsymbol{\sigma}(M(\underline{\mathbf{T}})_{:,:,i})$ denotes the singular value vector of the i-th frontal slice in the transformed domain.

[1] https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

II. *Intra-frequency low-rankness:* Within each active frequency component, the number of non-zero singular values is constrained to at most r. This condition ensures a low-rank structure for each frequency slice ($\forall i \in [m]$): $\sum_{j=1}^{\min\{d_1,d_2\}} \mathbb{I}\left(\sigma_j(M(\underline{\mathbf{T}})_{:,:,i}) \neq 0\right) \leq r$, where $\sigma_j(M(\underline{\mathbf{T}})_{:,:,i})$ denotes the j-th singular value of the i-th frontal slice of $M(\underline{\mathbf{T}})$.

This definition captures a strict form of dual-level structure by simultaneously enforcing sparsity across frequencies and low-rankness within each active frequency slice. While such hard constraints may be too restrictive in practical scenarios, especially where spectral contributions decay gradually, they provide a clear conceptual framework to motivate and analyze the more flexible soft regularization.

Soft Dual Spectral Sparsity. While the hard dual spectral sparsity model provides a clean conceptual foundation, its strict assumption of exact sparsity and fixed-rank constraints is often impractical in real-world scenarios. In many cases, singular values decay gradually rather than drop abruptly to zero, and the true number of active frequency components may be ambiguous or noise-sensitive. To overcome these limitations, we introduce a soft relaxation that allows for approximate sparsity and low-rankness in a continuous manner. Specifically, we propose the ℓ_p-Schatten-q quasi-norm, which relaxes the hard dual-sparsity constraints into a soft dual spectral sparsity framework.

Definition 3.2 (Tensor ℓ_p-Schatten-q quasi-norm)**.** For a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{d_1 \times d_2 \times m}$, we define its tensor ℓ_p-Schatten-q quasi-norm (abbreviated as $\ell_p(S_q)$-norm) as:

$$\|\underline{\mathbf{T}}\|_{\ell_p(S_q)} := \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{d_1 \wedge d_2} \sigma_j(M(\underline{\mathbf{T}})_{:,:,i})^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}, \tag{3}$$

where the exponents $(p, q) \in (0, 1]^2$.

In this quasi-norm, p governs the inter-frequency sparsity by promoting a group-wise regularization across frequency components, effectively highlighting significant groups while suppressing others. Simultaneously, q controls the intra-frequency low-rankness by encouraging sparsity in the singular values within each frequency slice, thereby modeling the intrinsic low-rank structure of the data. This soft dual spectral sparsity framework provides a unified yet versatile approach to address the hierarchical complexity of tensor data.

The ℓ_p-Schatten-q quasi-norm encompasses several existing regularization methods: it recovers TNN when $(p, q) = (1, 1)$[20], approximates the average rank as $(p, q) \to (1, 0)$[31], and reduces to the tensor Schatten-q norm when $p = q$ [12], thereby offering greater modeling flexibility. *Despite generalizing these regularizers, it fundamentally differs by jointly enforcing global frequency sparsity and local spectral low-rankness*.

While TNN applies uniform regularization across all singular values, the ℓ_p-Schatten-q quasi-norm introduces dual spectral sparsity control, modeling both inter-frequency sparsity through the ℓ_p-quasi-norm and intra-frequency low-rankness via the Schatten-q quasi-norm. This dual-level flexibility makes the proposed framework particularly well-suited for hierarchical and multi-scale data, where dependencies and sparsity exhibit layered structures. By bridging the gap between element-wise sparsity (as in TNN) and structured group sparsity, the ℓ_p-Schatten-q quasi-norm offers a more expressive and adaptable approach, enabling precise control over structural patterns in modern tensor-based analysis and recovery tasks.

4 Theory of Dual Spectral Sparse Tensor Estimation

This section develops the theoretical foundations of tensor estimation with dual spectral sparsity structures (**RQ2**).

Challenges. Dual spectral sparsity, combining inter-frequency sparsity with intra-frequency low-rankness, leads to *a globally coupled structure* that fundamentally differs from classical decoupled models like TNN. The ℓ_p-Schatten-q quasi-norm imposes interdependent constraints across frequency slices, resulting in a highly non-convex parameter space with nested sparsity patterns. This coupling prohibits slice-wise decomposition and complicates the use of standard tools. Accurately characterizing the estimation complexity demands novel extensions of covering numbers and metric entropy that jointly capture discrete sparsity and continuous low-rank structure.

To understand the statistical limits of learning under dual spectral sparsity, we analyze a simplified but representative model: the Gaussian location model, where the observed tensor is corrupted by additive noise. This setting preserves the core structural properties—inter-frequency sparsity and intra-frequency low-rankness—while avoiding complications unrelated to sparsity itself. Within this framework, we define structured parameter spaces that capture hard and soft variants of dual spectral sparsity, and establish sharp minimax lower and upper bounds under each. These results reveal how the joint effects of frequency selection and within-slice spectral decay determine the fundamental estimation limits, and provide theoretical justification for our proposed regularization.

4.1 Gaussian Location Model

Consider the Gaussian location model (GLM) [14], where n independent noisy realizations of the target tensor $\underline{\mathbf{L}}^* \in \mathbb{R}^{d_1 \times d_2 \times m}$ are observed as:

$$\underline{\mathbf{Y}}_i = \underline{\mathbf{L}}^* + \underline{\mathbf{E}}_i, \quad i \in [n], \tag{4}$$

where $\underline{\mathbf{Y}}_i \in \mathbb{R}^{d_1 \times d_2 \times m}$ is the observed tensor, $\underline{\mathbf{L}}^*$ represents the ground truth tensor of interest, and $\underline{\mathbf{E}}_i \in \mathbb{R}^{d_1 \times d_2 \times m}$ denotes the noise tensor with entries independently drawn from $\mathcal{N}(0, \sigma^2)$. The parameter σ characterizes the noise level. To simplify the analysis, we consider the sample mean of observations $\bar{\underline{\mathbf{Y}}} = n^{-1} \sum_{i=1}^n \underline{\mathbf{Y}}_i = \underline{\mathbf{L}}^* + \bar{\underline{\mathbf{E}}}$, where $\bar{\underline{\mathbf{E}}} = n^{-1} \sum_{i=1}^n \underline{\mathbf{E}}_i$ is the aggregated noise tensor with entries independently distributed as $\mathcal{N}(0, \sigma^2/n)$. The goal is to estimate the ground truth tensor $\underline{\mathbf{L}}^*$ based on the noisy observations $\{\underline{\mathbf{Y}}_i\}_{i=1}^n$. In particular, we aim to recover $\underline{\mathbf{L}}^*$ under dual spectral sparsity assumptions.

Remark 4.1. We adopt the Gaussian location model to isolate the core effects of dual spectral sparsity and the ℓ_p-Schatten-q regularization, avoiding additional complications from design tensors or sampling operators in tensor regression [35, 29, 24]. This simplified setting enables cleaner analysis and yields insights that extend naturally to regression problems under standard conditions such as RIP [35] or RSC [29, 24, 22].

Dual Spectral Sparsity Assumptions. We consider three distinct sparsity models for $\underline{\mathbf{L}}^*$:

A1. *Hard dual spectral sparsity*: Let $\underline{\mathbf{L}}^*$ belong to the parameter space

$$\mathsf{T}_{0,0}(s, r) = \{\underline{\mathbf{L}} : \text{at most } s \text{ active frequency slices, each of rank at most } r\}. \tag{5}$$

This model enforces exact inter-frequency sparsity and intra-frequency low-rankness.

A2. *Hard frequency sparsity and soft rank constraint (hard–soft sparsity)*: Let $\underline{\mathbf{L}}^*$ lie in

$$\mathsf{T}_{0,q}(s, R) = \left\{ \underline{\mathbf{L}} : |\{i : M(\underline{\mathbf{L}})_{:,:,i} \neq \mathbf{0}\}| \leq s, \ \|M(\underline{\mathbf{L}})_{:,:,i}\|_{S_q}^q \leq R, \ \forall i \in [m] \right\}. \tag{6}$$

This space imposes hard inter-frequency sparsity and soft Schatten-q constraints within each active slice.

A3. *Soft dual spectral sparsity*: Let $\underline{\mathbf{L}}^*$ belong to the parameter space

$$\mathsf{T}_{p,q}(R) = \left\{ \underline{\mathbf{L}} : \|\underline{\mathbf{L}}\|_{\ell_p(S_q)}^p \leq R \right\}. \tag{7}$$

Here, p promotes inter-frequency sparsity and q controls intra-frequency low-rankness via spectral decay; R specifies the quasi-norm ball radius.

These parameter spaces offer different views on structured tensor estimation: the *hard sparsity* model enforces strict thresholds, the *hard–soft model* balances structure with adaptability, and the *fully soft model* captures gradual spectral decay. Our goal is to estimate $\underline{\mathbf{L}}^*$ and derive minimax bounds under these assumptions.

4.2 Minimax Risk over Dual-level Sparse Structures

A key theoretical question in high-dimensional tensor estimation is: *What are the fundamental limits for recovering a tensor with dual spectral sparsity from noisy observations?* To address this, we establish minimax lower and upper bounds that characterize the best possible estimation accuracy achievable by any estimator under dual spectral sparsity assumptions.

$$\mathfrak{M}(\mathsf{T}) = \inf_{\hat{\underline{\mathbf{L}}}} \sup_{\underline{\mathbf{L}}^* \in \mathsf{T}} \mathbb{E}\left[\|\hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}^*\|_{\mathrm{F}}^2 \right], \tag{8}$$

where T is the parameter space. Following [18, 19], we consider $d_1 = d_2 = d$ for simplicity.

Theorem 4.2 (Minimax Bounds)**.** *The minimax risk under dual spectral sparsity satisfies the following bounds under certain conditions*[2]*:*

I. *Hard constraints on both frequency sparsity and per-slice low-rankness:*

$$\mathfrak{M}(\mathbf{T}_{0,0}(s,r)) \asymp \frac{\sigma^2}{n}\left(s\log\frac{em}{s} + srd\right).$$

II. *Hard frequency sparsity with soft intra-slice Schatten-q constraints:*

$$\mathfrak{M}(\mathbf{T}_{0,q}(s,R)) \asymp \frac{\sigma^2}{n} s\log\frac{em}{s} + sR\left(\frac{\sigma^2}{n}d\right)^{1-\frac{q}{2}}.$$

III. *Soft $\ell_p(S_q)$ constraints over both frequency and rank dimensions:*

$$\mathfrak{M}(\mathbf{T}_{p,q}(R)) \asymp \begin{cases} R\left(\frac{\sigma^2 n}{d}\right)^{\frac{p-2}{2}} + R\left(\frac{\sigma^2 n}{\log m}\right)^{\frac{p-2}{2}}, & p > q, \\ R^{\frac{q}{p}}\left(\frac{\sigma^2 n}{d}\right)^{\frac{q-2}{2}} + R\left(\frac{\sigma^2 n}{\log m}\right)^{\frac{p-2}{2}}, & p \le q,\ m > d^2, \\ R^{\frac{q}{p}}\left(\frac{\sigma^2 n}{d}\right)^{\frac{q-2}{2}}, & p \le q,\ m \le d^2. \end{cases}$$

Theorem 4.2 establishes the fundamental limits of estimation accuracy under different dual spectral sparsity structures. The minimax risk quantifies the worst-case squared Frobenius norm error that any estimator must incur when recovering a structured tensor from noisy observations. The results reveal the intricate balance between inter-frequency sparsity and intra-frequency low-rankness, showing how these factors jointly govern estimation complexity:

I. In the *hard sparsity* case, the estimation error consists of two terms: (i) $s\log(em/s)$, which reflects the difficulty of selecting s active frequency components, and (ii) srd, which characterizes the challenge of estimating rank-r matrices within each component.

II. In the *hard-soft sparsity* setting, the second term adapts to $sR(n^{-1}d)^{1-q/2}$, incorporating a smoother spectral decay controlled by q. Smaller q values impose stronger low-rank constraints, effectively reducing estimation complexity by promoting more aggressive rank sparsity.

III. In the *fully soft sparsity* scenario, where both inter-frequency sparsity and intra-frequency rank constraints are relaxed, the minimax risk follows distinct scaling behaviors across regimes. When $p > q$, the error rate is dominated by ℓ_p sparsity, with S_q low-rankness playing a minor role. For $p \le q$ and $m \ge d^2$, both the ℓ_p-ball and S_q-ball influence the estimation error, demonstrating an interplay between structured sparsity and low-rank regularization. When $m \le d^2$, the error rate is dictated by S_q, making it independent of m, emphasizing the fundamental role of rank constraints in this regime.

5 Optimization for Dual Spectral Sparse Tensor Estimation

Efficiently solving tensor estimation problems with dual spectral sparsity (**RQ3**) is key to leveraging the proposed ℓ_p-Schatten-q quasi-norm in practice. However, this task presents substantial challenges due to the non-convexity and coupled structure of this regularization.

Challenges. Even in the vector setting, optimizing dual-level sparse structures is notoriously difficult due to the combination of *non-convexity* and *structural coupling* [7, 15]. In our tensor case, these challenges are further compounded by the need to simultaneously enforce inter-frequency sparsity and intra-frequency low-rankness. Most existing tensor optimization methods either treat frequency components independently or impose low-rank constraints without spectral sparsity considerations, making them ill-suited for the proposed dual-spectral regularization. The ℓ_p-Schatten-q quasi-norm is non-convex whenever $p, q \in (0, 1]$, ruling out standard convex optimization techniques and necessitating a structure-aware, non-convex optimization strategy.

To address these difficulties, our approach is naturally motivated by the structural properties of the problem. We adopt a *proximal update scheme* that takes advantage of the separability of the

[2] The conditions in each setting are provided in the appendix.

transform-domain representation $M(\underline{\mathbf{L}})$, allowing frequency-wise updates, along with an iterative reweighting strategy that facilitates optimization in the presence of non-convex regularization.

Proximal Operator Formulation. To handle the non-convex ℓ_p-Schatten-q regularization, we adopt a proximal update scheme that enforces dual spectral sparsity while remaining computationally efficient. Specifically, at iteration t, the update is given by solving:

$$\underline{\mathbf{L}}^{t+1} \in \arg\min_{\underline{\mathbf{L}}} \frac{1}{2}\|\underline{\mathbf{L}} - \underline{\mathbf{Z}}\|_{\mathrm{F}}^2 + \lambda \sum\nolimits_{k=1}^{m} \|M(\underline{\mathbf{L}})_{:,:,k}\|_{S_q}^{p/q}, \tag{9}$$

where $\underline{\mathbf{Z}}$ denotes the intermediate variable aggregating previous updates and gradient information. Since the transform $M(\cdot)$ allows slice-wise decomposition [10], Problem (9) reduces to m subproblems over frequency components $k \in [m]$:

$$\min_{\mathbf{A}_k} \ \frac{1}{2} \|\mathbf{A}_k - M(\underline{\mathbf{Z}})_{:,:,k}\|_{\mathrm{F}}^2 + \lambda \|\mathbf{A}_k\|_{S_q}^{p/q}, \tag{10}$$

where $\mathbf{A}_k := M(\underline{\mathbf{L}})_{:,:,k}$ denotes the k-th frontal slice of the transformed tensor $M(\underline{\mathbf{L}})$. Problem (10) is difficult due to the non-convexity and lack of smoothness of the Schatten-q quasi-norm, which admits no closed-form or standard proximal solution in general.

To efficiently approximate Problem (10), we adopt a reweighted $\ell_{1/2}$-surrogate for $\|\mathbf{A}_k\|_{S_q}^{p/q}$ based on singular values:

$$\sum\nolimits_{i=1}^{d} w_{i,k} \cdot \sigma_i(\mathbf{A}_k)^{1/2}, \tag{11}$$

with weights defined as $w_{i,k} = \big(\sum_{j=1}^{d} \varsigma_{j,k}^{q} + \epsilon\big)^{p/q-1} \cdot \big(\varsigma_{i,k}^{1/2} + \epsilon\big)^{2q-1}$, where ϵ is a small regularization constant and $\varsigma_{j,k} := \sigma_j(M(\underline{\mathbf{L}}^t)_{:,:,k})$ are the singular values from the previous iterate. The update for each singular value then becomes a soft-thresholding step:

$$\sigma_i^{(t+1)}(M(\underline{\mathbf{L}})_{:,:,k}) = \mathcal{S}_{\lambda w_{i,k}}^{\ell_{1/2}}\left(\sigma_i(M(\underline{\mathbf{Z}})_{:,:,k})\right), \tag{12}$$

where $\mathcal{S}^{\ell_{1/2}}$ is the proximal operator for the $\ell_{1/2}$-norm (see Appendix for closed-form expression).

After singular value shrinkage, we reconstruct each slice $M(\underline{\mathbf{L}}^{t+1})_{:,:,k} = \mathbf{U}_k \cdot \mathrm{diag}(\boldsymbol{\sigma}^{(t+1)}) \cdot \mathbf{V}_k^\top$, where $\mathbf{U}_k$ and $\mathbf{V}_k$ are from the SVD of $M(\underline{\mathbf{Z}})_{:,:,k}$. Finally, applying the inverse transform yields the updated tensor $\underline{\mathbf{L}}^{t+1}$ in the original domain.

6 Experiments

Having established the theoretical foundations and algorithmic framework, we now evaluate the empirical performance of the proposed ℓ_p-Schatten-q quasi-norm in tensor estimation tasks. We conduct extensive experiments on three types of remote sensing data to demonstrate its effectiveness in noisy tensor completion tasks.

Experimental Setup. We consider the noisy tensor completion which involves reconstructing a tensor from noisy incomplete observations. Given a clean tensor $\underline{\mathbf{L}}$ of size $d_1 \times d_2 \times d_3$, we introduce *i.i.d.* Gaussian noise with standard deviation $\sigma = c\sigma_0$, where $c = 0.05$ and $\sigma_0 = \|\underline{\mathbf{L}}\|_{\mathrm{F}} / \sqrt{d_1 d_2 d_3}$. A uniform sampling strategy is applied with sampling ratios $p \in \{0.05, 0.1, 0.15\}$, meaning that 95%, 90%, and 85% of the entries are missing, respectively. Each setting is tested over 10 trials, and the averaged PSNR (dB) and SSIM values are reported. To benchmark our method, we compare the proposed $\ell_p(S_q)$-quasi-norm against several low-rank regularizers, including matrix nuclear norm (NN) [2], Tucker-based tensor nuclear norm (SNN) [16], TNN-DFT [37], TNN-DCT [20], tensor k-Support norm (k-Supp) ($k = 2$) [29], tensor ℓ_{1-2}-norm (ℓ_{1-2}) [26], tensor Schatten-p-norm ($p = 1/2$) [12]. In our implementation, we set the sparsity parameters[3] to $(p, q) = (0.8961, 0.8966)$ and employ the Discrete Cosine Transform (DCT) as the transform operator $M(\cdot)$. Details of the experiments are given in the appendix.

[3]We first performed a coarse grid search over $p, q \in \{0.1, 0.2, \ldots, 1.0\}$ and observed consistent performance peaks near $p = q = 0.9$. We then manually fine-tuned within $[0.88, 0.92]$ based on PSNR, selecting $(p, q) = (0.8961, 0.8966)$ as the best-performing pair.

Table 1: Results for noisy tensor completion on remote sensing datasets are shown below. The best result in each case is highlighted in **bold**, while the second-best is $\underline{\text{underlined}}$.

Dataset	SR	Metric	NN	SNN	TNN-DFT	TNN-DCT	k-Supp	ℓ_{1-2}	Schatten-$1/2$	$\ell_p(S_q)$ (proposed)
SalinasA	5%	PSNR	15.21	20.79	22.55	$\underline{26.52}$	22.58	22.21	22.45	**28.43**
		SSIM	0.2594	**0.7547**	0.5667	0.7384	0.5689	0.5524	0.4474	$\underline{0.7374}$
	10%	PSNR	20.62	25.56	25.72	$\underline{29.61}$	25.89	26.14	25.86	**31.81**
		SSIM	0.4775	0.8284	0.7027	$\underline{0.8403}$	0.7231	0.7197	0.6058	**0.8484**
	15%	PSNR	23.09	27.99	28.06	$\underline{31.32}$	28.09	28.13	26.98	**33.23**
		SSIM	0.5643	0.8622	0.7804	$\underline{0.8798}$	0.7810	0.7795	0.6505	**0.8830**
IndianPines	5%	PSNR	20.44	22.01	25.68	$\underline{26.26}$	25.70	25.73	24.68	**27.05**
		SSIM	0.3895	0.6359	0.6293	$\underline{0.6727}$	0.6289	0.6316	0.5361	**0.6740**
	10%	PSNR	22.23	24.94	27.45	$\underline{28.40}$	27.48	27.52	25.72	**28.92**
		SSIM	0.4836	0.7171	0.7226	**0.7744**	0.7219	0.7249	0.5991	$\underline{0.7617}$
	15%	PSNR	23.52	26.61	28.54	$\underline{29.52}$	28.53	28.63	26.24	**29.89**
		SSIM	0.5438	0.7668	0.7713	**0.8177**	0.7709	0.7741	0.6258	$\underline{0.7997}$
Cloth	5%	PSNR	20.10	20.95	25.00	$\underline{26.09}$	25.08	25.09	24.96	**26.99**
		SSIM	0.3762	0.5096	0.6773	$\underline{0.7283}$	0.6792	0.6793	0.6305	**0.7422**
	10%	PSNR	21.14	22.72	28.00	$\underline{29.24}$	28.12	28.14	27.98	**30.63**
		SSIM	0.4341	0.5983	0.8132	$\underline{0.8540}$	0.8143	0.8163	0.7668	**0.8658**
	15%	PSNR	22.05	24.18	30.03	$\underline{31.36}$	30.08	30.11	29.50	**32.71**
		SSIM	0.4889	0.6783	0.8722	$\underline{0.9054}$	0.8727	0.8733	0.8153	**0.9090**
Hair	5%	PSNR	25.33	30.09	33.16	$\underline{35.31}$	33.19	33.27	33.43	**36.95**
		SSIM	0.7147	0.8631	0.8917	**0.9248**	0.8921	0.8919	0.8240	$\underline{0.9196}$
	10%	PSNR	29.52	33.35	36.22	$\underline{38.18}$	36.17	36.30	35.69	**39.91**
		SSIM	0.8008	0.9122	0.9292	**0.9535**	0.9286	0.9296	0.8640	$\underline{0.9517}$
	15%	PSNR	31.12	35.24	38.00	$\underline{39.88}$	37.91	38.07	36.46	**41.52**
		SSIM	0.8364	0.9336	0.9449	**0.9650**	0.9442	0.9448	0.8735	$\underline{0.9641}$
JellyBeans	5%	PSNR	16.33	18.21	25.43	$\underline{26.47}$	25.38	25.62	25.39	**27.91**
		SSIM	0.2397	0.4942	0.6726	**0.7223**	0.6714	0.6733	0.5504	$\underline{0.7115}$
	10%	PSNR	18.12	22.11	28.50	$\underline{30.14}$	28.47	28.67	28.41	**31.95**
		SSIM	0.3169	0.6629	0.7900	**0.8518**	0.7902	0.7932	0.6905	$\underline{0.8486}$
	15%	PSNR	19.92	24.67	30.51	$\underline{32.33}$	30.52	30.61	29.96	**33.97**
		SSIM	0.4053	0.7592	0.8489	**0.9030**	0.8504	0.8499	0.7516	$\underline{0.8980}$
OSU Thermal	5%	PSNR	13.19	15.83	28.06	27.99	28.01	$\underline{28.19}$	28.11	**30.06**
		SSIM	0.1848	0.4759	0.8584	$\underline{0.8707}$	0.8579	0.8603	0.7928	**0.8759**
	10%	PSNR	14.67	19.75	31.30	$\underline{31.62}$	31.28	31.60	30.51	**33.67**
		SSIM	0.2509	0.6594	0.9151	**0.9326**	0.9147	0.9168	0.8358	$\underline{0.9272}$
	15%	PSNR	16.27	22.52	33.02	$\underline{33.51}$	33.05	33.11	30.99	**35.09**
		SSIM	0.3273	0.7621	0.9315	**0.9509**	0.9321	0.9318	0.8373	$\underline{0.9404}$

Datasets. We validate our approach on three categories of remote sensing data. First, for hyperspectral images, we employ the corrected *Indian Pines* and *Salinas A* datasets from the AVIRIS sensor, containing 200 and 204 spectral bands respectively. Due to computational considerations, we utilize the first 30 bands in our experiments. Second, we evaluate on multispectral images from the Columbia MSI Database, including *Cloth*, *Hair*, and *Jelly Beans*, each with dimensions $512 \times 512 \times 31$ and normalized intensity values in [0,1]. Finally, for thermal imaging, we use sequences from the *OSU Thermal Database*, specifically the first 30 frames of Sequence 1, forming a tensor of size $320 \times 240 \times 30$.

Results and Analysis. Table 1 summarizes the PSNR and SSIM results across different missing rates. The proposed $\ell_p(S_q)$-quasi-norm achieves the highest PSNR, demonstrating its effectiveness in preserving spectral information. Its SSIM results rank among the top two, indicating that our approach better retains structural integrity compared to competing methods. These experimental results demonstrate the effectiveness of the proposed ℓ_p-Schatten-q quasi-norm in robust tensor recovery, showing how characterizing dual spectral sparsity structures in transformed domains benefits tensor reconstruction performance.

7 Conclusion

This paper identifies and formalizes a coupled spectral structure within the t-SVD framework, where inter-frequency sparsity coexists with intra-frequency low-rankness. To capture this structure, we propose a unified modeling approach based on the ℓ_p-Schatten-q quasi-norm, which enables separate control over spectral sparsity at different levels and generalizes existing tensor norms. We provide sharp minimax guarantees under both hard and soft sparsity regimes, and develop an efficient proximal algorithm tailored to this setting. Experimental results demonstrate the practical potential of the proposed approach for structured tensor recovery.

Limitation. To highlight the fundamental properties of the proposed ℓ_p-Schatten-q quasi-norm, our analysis employs several simplifications, including Gaussian location model and idealized sparsity patterns. While our optimization algorithm shows promising empirical performance, its theoretical convergence properties remain to be established. These theoretical and algorithmic limitations suggest important directions for future research.

References

[1] G. Bergqvist and E. G. Larsson. The higher-order singular value decomposition: Theory and an application [lecture notes]. *IEEE signal processing magazine*, 27(3):151–154, 2010.

[2] E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[3] J. D. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.

[4] A. Cichocki, N. Lee, I. Oseledets, A. H. Phan, Q. Zhao, and D. P. Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations & Trends ® in Machine Learning*, 9(4-5):249–429, 2016.

[5] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

[6] J. Hou, F. Zhang, H. Qiu, J. Wang, Y. Wang, and D. Meng. Robust low-tubal-rank tensor recovery from binary measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[7] Y. Hu, C. Li, K. Meng, J. Qin, and X. Yang. Group sparse optimization via $\ell_{p,q}$ regularization. *Journal of Machine Learning Research*, 18(30):1–52, 2017.

[8] M. Imaizumi, T. Maehara, and K. Hayashi. On tensor train rank minimization: Statistical efficiency and scalable algorithm. In *Advances in Neural Information Processing Systems*, pages 3930–3939, 2017.

[9] E. Kernfeld, M. Kilmer, and S. Aeron. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015.

[10] M. E. Kilmer, K. Braman, et al. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J MATRIX ANAL A*, 34(1):148–172, 2013.

[11] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[12] H. Kong, X. Xie, and Z. Lin. t-Schatten-p norm for low-rank tensor recovery. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1405–1419, 2018.

[13] C. Li, W. He, L. Yuan, Z. Sun, and Q. Zhao. Guaranteed matrix completion under multiple linear transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11136–11145, 2019.

[14] Z. Li, Y. Zhang, and J. Yin. Estimating double sparse structures over $\ell_u(\ell_q)$-balls: Minimax rates and phase transition. *IEEE Transactions on Information Theory*, 2024.

[15] R. Lin, S. Chen, H. Feng, and Y. Liu. Computing the proximal operator of the $\ell_{p,q}$-norm for group sparsity. *arXiv preprint arXiv:2409.14156*, 2024.

[16] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2013.

[17] X. Liu, S. Aeron, V. Aggarwal, and X. Wang. Low-tubal-rank tensor completion using alternating minimization. *IEEE Transactions on Information Theory*, 66(3):1714–1737, 2020.

[18] C. Lu. Transforms based tensor robust PCA: Corrupted low-rank tensors recovery via convex optimization. In *ICCV*, pages 1145–1152, 2021.

[19] C. Lu, J. Feng, W. Liu, Z. Lin, S. Yan, et al. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[20] C. Lu, X. Peng, and Y. Wei. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *CVPR*, pages 5996–6004, 2019.

[21] C. D. Martin, R. Shafer, and B. Larue. An order-p tensor factorization with applications in imaging. *SIAM Journal on Scientific Computing*, 35(1), 2013.

[22] S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13:1665–1697, 2012.

[23] I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[24] Y. Qiu, G. Zhou, A. Wang, Q. Zhao, and S. Xie. Balanced unfolding induced tensor nuclear norms for high-order tensor completion. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[25] G. Song, M. K. Ng, and X. Zhang. Robust tensor completion using transformed tensor singular value decomposition. *NUMER LINEAR ALGEBR*, 27(3):e2299, 2020.

[26] Z. Tan, L. Huang, H. Cai, and Y. Lou. Non-convex approaches for low-rank tensor completion under tubal sampling. In *ICASSP*, pages 1–5. IEEE, 2023.

[27] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[28] A. Wang, C. Li, M. Bai, Z. Jin, G. Zhou, and Q. Zhao. Transformed low-rank parameterization can help robust generalization for tensor neural networks. *NeurIPS*, 36, 2023.

[29] A. Wang, G. Zhou, Z. Jin, and Q. Zhao. Tensor recovery via $*_L$-spectral k-support norm. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):522–534, 2021.

[30] H. Wang, J. Yang, X. Yu, Y. Zhang, J. Qian, and J. Wang. Tensor-flamingo unravels the complexity of single-cell spatial architectures of genomes at high-resolution. *Nature Communications*, 16(1):3435, 2025.

[31] Z. Wang, J. Dong, X. Liu, and X. Zeng. Low-rank tensor completion by approximating the tensor average rank. In *ICCV*, pages 4612–4620, 2021.

[32] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126(11):1157–1179, 2018.

[33] Y. Xu, R. Hao, W. Yin, and Z. Su. Parallel matrix factorization for low-rank tensor completion. *Inverse Problems and Imaging*, 9(2):601–624, 2015.

[34] M. Yuan and C. H. Zhang. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1–38, 2016.

[35] F. Zhang, W. Wang, J. Huang, J. Wang, and Y. Wang. RIP-based performance guarantee for low-tubal-rank tensor recovery. *Journal of Computational and Applied Mathematics*, 374:112767, 2020.

[36] X. Zhang and M. K.-P. Ng. Low rank tensor completion with poisson observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[37] Z. Zhang and S. Aeron. Exact tensor completion using t-svd. *IEEE Transactions on Signal Processing*, 65(6):1511–1526, 2017.

[38] Z. Zhang, G. Ely, S. Aeron, et al. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *CVPR*, pages 3842–3849, 2014.

[39] P. Zhou and J. Feng. Outlier-robust tensor PCA. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the central focus of the paper—the modeling, theoretical analysis, and algorithmic solution for dual spectral sparsity in tensors—which directly correspond to the three core contributions developed in the main body.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The paper includes a dedicated discussion on limitations in the conclusion section, noting simplifications such as the use of the Gaussian location model and the lack of theoretical convergence guarantees for the proposed non-convex optimization algorithm.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: For each theoretical result, the paper states all necessary assumptions explicitly and provides complete proofs in the appendix. The analysis includes both lower and upper minimax bounds under different dual-sparsity regimes, supported by standard and extended techniques such as entropy numbers and packing arguments.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides sufficient information to reproduce the main experimental results, including dataset descriptions, sampling settings, noise levels, baseline configurations, and implementation details of the proposed algorithm. Additional algorithmic and parameter details are included in the appendix to support reproducibility.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The authors have uploaded the code as part of the supplementary material, along with sufficient implementation details and instructions to reproduce the main experimental results. All datasets used are publicly available and clearly specified.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).

- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies all necessary details, including sampling ratios, noise levels, hyperparameter settings, evaluation metrics (PSNR and SSIM), and implementation choices. Additional settings such as initialization and convergence criteria are provided in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports averaged performance over multiple random trials for each experimental setting. While explicit error bars are not shown in tables, standard practice is followed by reporting stable metrics (PSNR and SSIM) under fixed noise and sampling ratios, ensuring reliable comparative evaluation.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).

- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The appendix provides information on the computational environment used for the experiments, including hardware specifications such as CPU type and memory.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research does not involve human subjects, personally identifiable information, sensitive data, or potentially harmful applications. It focuses on theoretical modeling and algorithmic development for tensor estimation, aligning fully with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The paper focuses solely on the theoretical modeling and optimization of the tensor recovery algorithm, without involving any societal issues, and therefore does not require discussion of societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This work does not involve the release of high-risk models or datasets. It focuses on a general-purpose tensor estimation framework using publicly available data, and does not include pretrained models or components with significant misuse potential.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external datasets and baseline implementations used in the paper are publicly available and properly cited. Their licenses and terms of use have been respected, and relevant references are provided in the main text and appendix.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The paper introduces a new optimization algorithm and associated code, which are included in the supplementary material. The code is well documented with clear instructions, comments, and reproducibility guidelines to support independent use and verification.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This work does not involve any human subjects or crowdsourcing experiments.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This study does not involve human participants and therefore does not require IRB approval or risk disclosure.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: This work does not use large language models (LLMs) as part of its core methodology or experimental pipeline.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.