Towards a Geometric Understanding of Tensor Learning via the t-Product

Andong Wang
RIKEN AIP
andong.wang@riken.jp

Yuning Qiu
RIKEN AIP
yuning.qiu@riken.jp

Haonan Huang
RIKEN AIP
haonan.huang@riken.jp

Zhong Jin
China University of Petroleum-Beijing at Karamay
zhongjin@cupk.edu.cn

Guoxu Zhou
Guangdong University of Technology
gx.zhou@gdut.edu.cn

Qibin Zhao*
RIKEN AIP
qibin.zhao@riken.jp

Abstract

Despite the growing success of transform-based tensor models such as the t-product, their underlying geometric principles remain poorly understood. Classical differential geometry, built on real-valued function spaces, is not well suited to capture the algebraic and spectral structure induced by transform-based tensor operations. In this work, we take an initial step toward a geometric framework for tensors equipped with tube-wise multiplication via orthogonal transforms. We introduce the notion of smooth t-manifolds, defined as topological spaces locally modeled on structured tensor modules over a commutative t-scalar ring. This formulation enables transform-consistent definitions of geometric objects, including metrics, gradients, Laplacians, and geodesics, thereby bridging discrete and continuous tensor settings within a unified algebraic-geometric perspective. On this basis, we develop a statistical procedure for testing whether tensor data lie near a low-dimensional t-manifold, and provide nonasymptotic guarantees for manifold fitting under noise. We further establish approximation bounds for tensor neural networks that learn smooth functions over t-manifolds, with generalization rates determined by intrinsic geometric complexity. This framework offers a theoretical foundation for geometry-aware learning in structured tensor spaces and supports the development of models that align with transform-based tensor representations.

1 Introduction

Tensor-based modeling has emerged as a powerful paradigm for representing and analyzing structured data, with widespread applications in machine learning, computer vision, signal processing, and quantum AI [64, 65, 78, 53, 23, 39, 61, 62, 83]. Among these, the t-SVD framework [33, 32] introduces *a unique t-scalar representation*, where each mode-3 fiber (tube) is treated as an indivisible algebraic unit called a t-scalar, following a transform-based multiplication rule (see Section 2.1 for details). Built on this foundation, a 2D image can be viewed as a t-vector whose entries are t-scalars, while a 3D video can be interpreted as a t-matrix [79, 41, 33]. This representation enables low-rank

*Qibin Zhao is the corresponding author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

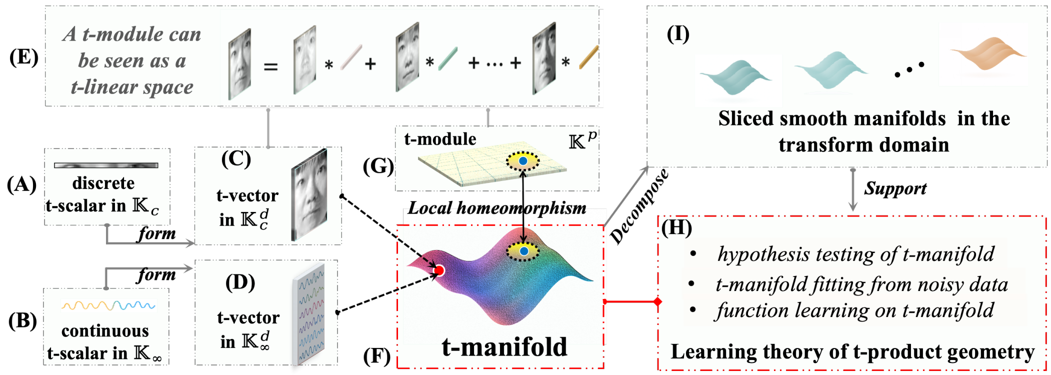

Figure 1: Overview of the proposed t-product geometric framework. (A, B) Discrete and continuous t-scalars in $\mathbb{K}_c$ and $\mathbb{K}_\infty$ form high-dimensional t-vectors (C, D) as elements of $\mathbb{K}_c^d$ or $\mathbb{K}_\infty^d$. (E) A t-module is a t-linear space spanned by t-vectors with coefficients in t-scalars. (F) A t-manifold is defined as a smooth space locally homeomorphic to a (finitely generated free) t-module (G), generalizing classical manifolds to the transform-based tensor setting. (I) In the transform domain, the t-manifold decomposes into multiple frequency-wise manifolds, which support spectral smoothness. (H) These constructions support the development of learning theory over t-manifolds, including hypothesis testing, fitting from noisy data, and smooth function learning with tensor neural networks.

modeling in the transform domain, effectively capturing structured dependencies such as spatial, spectral, and channel-wise correlations [32, 31, 83, 68, 27, 71].

A central concept in many t-SVD-based methods, such as Tensor Robust PCA [41] and Tensor Low-Rank Representation (TLRR) [83], is the structure that can be viewed as a *t-module* (see Section 2.1 and Fig. 1-E), which serves as an analog of a linear space defined over the t-product algebra [30, 6]. For instance, TLRR models utilize t-modules to express the self-representation property of tensor data [83, 68, 69, 30]. While effective, such linear structures are inherently limited in their ability to capture nonlinear interactions or encode higher-order geometric properties. Moreover, existing approaches are primarily formulated based on discrete tube structures [83, 68, 69, 30], typically corresponding to time-indexed sequences with finite sampling, such as video frames [83]. These methods lack a unified perspective that bridges discrete-time tubes [83] and continuous-time tube representations [67, 47, 66] within a coherent geometric framework.

Recent work has begun to explore geometric aspects of the t-product, including Grassmannian and Stiefel manifolds defined over t-scalars [45, 19]. However, these developments remain narrowly focused and lack a unified theoretical framework. In particular, a general theory of differential geometry that is compatible with the t-product and capable of defining smooth structures, Riemannian metrics, geodesics, and other geometric objects in a transform-consistent manner has yet to be established. At the same time, t-SVD-based models have demonstrated strong empirical performance in graph-structured learning tasks [13, 51, 24, 70], where graphs may be viewed as discrete samples from an underlying manifold. Yet, the intrinsic geometric structure of such tensorized data remains largely unexplored.

These observations point to an important yet underdeveloped direction: *developing a geometric framework for tensor learning that extends differential geometry to the algebra induced by the t-product*. Such a framework is essential not only for unifying discrete and continuous tensor modeling, but more importantly, for establishing a principled theoretical foundation for structure-aware learning in tensor settings. In this work, we make an initial attempt toward this goal by posing the following theoretical questions:

- **Q1**: *Can we construct a rigorous differential geometric framework over t-scalars that systematically unifies discrete and continuous tensor representations, and gives rise to well-defined geometric objects such as t-manifolds?*
- **Q2**: *Can such a framework support the theoretical analysis of smooth function learning on low-dimensional t-manifolds from high-dimensional tensor data?*

To address these questions, we propose a theoretical framework of *t-product geometry* (see Fig. 1) with the following contributions:

- **Geometry.** We introduce the concept of *t-manifolds*, a new class of smooth manifolds defined over the t-scalar ring, formalized via a sheaf-theoretic construction. This lays a principled foundation for defining transform-consistent geometric structures such as tangent spaces and Riemannian metrics (Section 2), directly addressing **Q1**.
- **Learning theory.** We develop hypothesis testing and fitting methods for identifying low-dimensional t-manifold structures from high-dimensional tensor data, and establish theoretical guarantees for learning smooth functions defined on such manifolds using tensor neural networks. These results offer a rigorous theoretical foundation for geometry-aware learning in the tensor setting (Sections 3.1–3.2), thereby addressing **Q2**.

Beyond the theoretical developments, we further illustrate the modeling potential of the proposed framework through a conceptual example in image modeling, where bidirectional structures based on the t-product formulation are utilized to enhance clustering and tensor recovery[2] (Section 4). To our knowledge, this work provides the first framework that formulates tensor learning within a differential-geometric space defined by the t-product algebra, bridging transform-based tensor analysis and smooth manifold theory. The appendix provides detailed related work, proofs, algorithms, and experiments.

2 Differential Geometry over the t-Product Algebra

To address **Q1**, we develop a differential geometric framework over the t-product algebra by introducing a new class of smooth manifolds, which we refer to as *t-manifolds*. These manifolds are constructed over the t-scalar ring and support transform-consistent notions of smoothness and locality. We begin by introducing the algebraic foundations necessary for this formulation.

2.1 Preliminaries on t-Scalar-Based Representation

Discrete t-scalars and the ring $\mathbb{K}_c$. A *discrete t-scalar* is a third-order tensor $\boldsymbol{a} \in \mathbb{R}^{1\times 1\times c}$, which may represent a row in a 2D image (with c columns, see Fig. 1-A) [83, 68] or the RGB channels of a pixel (when $c = 3$) [41]. We define the *t-product* between t-scalars as a multiplication on $\mathbb{R}^{1\times 1\times c}$ via an orthogonal transform $M \in \mathbb{R}^{c\times c}$ (e.g., DCT) [61, 62], letting $\boldsymbol{a} * \boldsymbol{b} := M^{-1}(M(\boldsymbol{a}) \odot M(\boldsymbol{b}))$, where $\odot$ is the Hadamard product. This turns the space into a commutative ring [6], denoted $\mathbb{K}_c := (\mathbb{R}^{1\times 1\times c}, +, *)$, with identity element $e := M^{-1}(\mathbf{1})$ where $\mathbf{1} \in \mathbb{R}^c$ is the all-ones vector.

Continuous t-scalars and the ring $\mathbb{K}_\infty$. In the continuous setting, a *continuous t-scalar* is a smooth function $f \in C^\infty(\mathbb{R})$, representing, e.g., a continuous-time signal (see Fig. 1-B) [67]. Let M be a unitary operator on $C^\infty(\mathbb{R})$ (e.g., a multiplication or translation operator). Then multiplication is defined as $f * g := M^{-1}((Mf)(Mg))$, and the ring[3] is denoted as $\mathbb{K}_\infty := (C^\infty(\mathbb{R}), +, *)$. For simplicity, we assume the existence of a unit element $e = M^{-1}(1)$.

T-vectors and t-module. A *t-vector* is a tuple of d t-scalars (see Figs. 1-C and D). In the discrete case, $\boldsymbol{p} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d) \in \mathbb{K}_c^d$ with each $\boldsymbol{p}_i \in \mathbb{R}^{1\times 1\times c}$; in the continuous case, $f = (f_1, \ldots, f_d) \in \mathbb{K}_\infty^d$. This allows us to represent structured objects such as image rows, multichannel signals, or time-varying slices [61, 83, 67]. The set $\mathbb{K}^d$ forms a free right module over the t-scalar ring $\mathbb{K} \in \{\mathbb{K}_c, \mathbb{K}_\infty\}$, supporting addition and scalar multiplication defined componentwise [6]. We refer to such a space as a *t-module*, which can be regarded as a natural generalization of a linear subspace, where the scalar field (e.g., $\mathbb{R}$ or $\mathbb{C}$) is replaced by the t-product algebra $\mathbb{K}$ (see Fig. 1-E). This t-module structure serves as the model space for defining local charts in our construction of t-manifolds (Definition 1).

[2] This example is intended to demonstrate the modeling perspective inspired by the t-product framework rather than to serve as an empirical benchmark for t-manifold learning.

[3] Alternatively, one may define $\mathbb{K}_\infty$ over $C(\mathbb{R})$ rather than $C^\infty(\mathbb{R})$, provided that the transform M preserves continuity and ensures closure of the induced multiplication in $C(\mathbb{R})$, and that the unit element $e = M^{-1}(1)$ exists. For instance, when M is a translation or a smooth multiplication operator, e is smooth and lies in $C^\infty(\mathbb{R})$. In contrast, if M is the Fourier transform, then $e = \delta(t)$ is a tempered distribution lying outside $C(\mathbb{R})$; this can be addressed by extending the scalar ring to the space of distributions $\mathcal{D}'(\mathbb{R})$ or by adjoining a formal identity.

2.2 Smooth t-Manifolds with Intrinsic Dimension p

We propose the notion of *t-manifolds* as a generalization of classical smooth manifolds (see Fig. 1-F), in which real-valued coordinates are replaced by structured t-scalars from a commutative ring $\mathbb{K} \in \{\mathbb{K}_c, \mathbb{K}_\infty\}$, and the local structure is modeled by a t-module (see Fig. 1-G). This provides a unified algebraic-geometric view of discrete and continuous tensor data.

Definition 1 (Smooth t-Manifold of Dimension p)**.** Let $\beta \in \mathbb{N}_{\geqslant 1}$. A topological space $\mathcal{M}$ is called a C^β-smooth t-manifold of intrinsic dimension p over $\mathbb{K}$ if the following conditions are satisfied:

(1) Topological structure. $\mathcal{M}$ is a Hausdorff and paracompact topological space.

(2) Local charts. There exists a locally finite atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ such that each U_α is open in $\mathcal{M}$, and each φ_α is a homeomorphism onto an open subset $V_\alpha \subseteq \mathbb{K}^p$, where $\mathbb{K}^p$ denotes the free $\mathbb{K}$-module of rank p (with $\mathbb{K} \in \{\mathbb{K}_c, \mathbb{K}_\infty\}$). The manifold $\mathcal{M}$ can be regarded as embedded in a higher-rank free $\mathbb{K}$-module $\mathbb{K}^d$ ($d \geqslant p$), equipped with the product topology—Euclidean if $\mathbb{K} = \mathbb{K}_c$ and Fréchet if $\mathbb{K} = \mathbb{K}_\infty$.

(3) Component-wise smoothness and independence. For any overlapping charts φ_α and φ_β with $U_\alpha \cap U_\beta \neq \varnothing$, the transition map $\varphi_\alpha \circ \varphi_\beta^{-1}$ is C^β-smooth in the transform domain, and acts independently[4] across frequency components. Specifically:

- *(Discrete case)* $\mathbb{K} = \mathbb{K}_c$: For each frequency index $k = 1, \ldots, c$, the k-th slice

  $$[M \circ \varphi_\alpha \circ \varphi_\beta^{-1} \circ M^{-1}]_k : \mathbb{R}^p \to \mathbb{R}^p$$

  is a C^β-smooth diffeomorphism, and different slices are mutually independent. Consequently, $M(\mathcal{M}) \cong \mathcal{M}^{(1)} \times \cdots \times \mathcal{M}^{(c)}$, where each $\mathcal{M}^{(k)}$ is a C^β-smooth p-dimensional manifold.

- *(Continuous case)* $\mathbb{K} = \mathbb{K}_\infty$: For each frequency parameter $t \in \mathbb{R}$,

  $$[M \circ \varphi_\alpha \circ \varphi_\beta^{-1} \circ M^{-1}](t) : \mathbb{R}^p \to \mathbb{R}^p$$

  is C^β-smooth, and the map $t \mapsto [M \circ \varphi_\alpha \circ \varphi_\beta^{-1} \circ M^{-1}](t)$ is continuous in the C^β-topology. Assuming component-wise independence across t, the transform-domain representation satisfies $M(\mathcal{M}) \cong \int_{\mathbb{R}}^{\oplus} \mathcal{M}^{(t)}\, dt$, where $\{\mathcal{M}^{(t)}\}_{t \in \mathbb{R}}$ forms a continuously parameterized family of C^β-smooth manifolds.

This definition generalizes classical smooth manifolds: when $\mathbb{K} = \mathbb{R}$ and $M = \mathrm{Id}$, a t-manifold reduces to an ordinary C^β-smooth manifold modeled on $\mathbb{R}^p$, with standard smooth transition maps. More generally, when $\mathbb{K} \in \{\mathbb{K}_c, \mathbb{K}_\infty\}$ and the product topology is imposed, the free module $\mathbb{K}^p$ becomes a trivial C^β-smooth t-manifold and serves as the local model space for general t-manifolds. Definition 1 thus extends the notion of smooth manifolds to the algebraic setting induced by the t-product, where smoothness is enforced in the transform domain to ensure compatibility with underlying tensor operations. This definition strictly adheres to the core philosophy of the t-SVD [33, 32, 6]: *by treating tubes as algebraic units and operating in the transform domain, it captures low-rank structures and coherent variations along spectral modes while ensuring frequency-wise independence among transform components.*

In data modeling, the intrinsic dimension p quantifies the local degrees of freedom required to parametrize the manifold $\mathcal{M}$, while the ambient dimension d corresponds to the number of t-vector coordinates (e.g., image rows or spatial locations). In most applications, we expect $d \gg p$, consistent with the *manifold hypothesis*–the assumption that structured tensor data concentrate near a low-dimensional t-manifold embedded in a high-dimensional t-vector space. This hypothesis will be theoretically examined in Section 3.1.

[4]Throughout this work, t-manifolds are defined under the assumption of *frequency-wise independence* in the transform domain, meaning that local smooth structures are specified separately for each frequency component. This assumption simplifies the definition of differential and geometric objects but rules out cross-frequency interactions that could model spectral coupling. Relaxing this independence assumption in manners like [63] would allow for more expressive geometric structures and constitutes an interesting direction for future research.

2.3 Differential Geometry on t-Manifolds

To develop a coherent geometric framework on t-manifolds, classical differential geometry needs to be extended in a manner that *respects the algebraic structure induced by the t-product*. Traditional geometric objects such as vector fields, differential forms, and Riemannian metrics are defined over real-valued functions. In contrast, t-manifolds are locally modeled on t-scalars drawn from a transform-based commutative ring $\mathbb{K}$ (e.g., $\mathbb{K}_c$ or $\mathbb{K}_\infty$), rather than $\mathbb{R}$. This shift leads to a basic issue: *since t-scalars are structured objects, such as discrete tubes or smooth functions, basic operations including directional derivatives and linear functionals cannot be defined by direct scalar-based analogies.* Instead, they must be formulated in a manner consistent with the underlying transform algebra.

To address this, we adopt a sheaf-theoretic formulation rooted in commutative algebra [7, 22, 20], which provides a unified language for organizing frequency-wise smooth functions and their differential relations. In this setting, all geometric objects are defined with respect to the structure sheaf $\mathcal{O}_\mathcal{M}$ of $\mathbb{K}$-smooth functions, ensuring transform-consistent local definitions even under frequency-wise independence. Tangent vector fields are realized as $\mathbb{K}$-linear derivations on $\mathcal{O}_\mathcal{M}$, generalizing directional derivatives to the setting of structured scalars, while differential 1-forms are defined as elements of the Kähler differential module [22] over $\mathcal{O}_\mathcal{M}$, satisfying the Leibniz rule and serving as duals to vector fields under evaluation.

We now introduce the algebraic structures underlying differential geometry on t-manifolds.

Algebraic structures for geometry on t-manifolds. We generalize classical differential constructions to the setting of $\mathbb{K}$-smooth functions using a sheaf-theoretic approach that is compatible with the transform-based algebra underlying the t-product.

Definition 2 ($\mathbb{K}$-Smooth Function)**.** Let U be an open subset of a t-manifold $\mathcal{M}$ and M be the transform defining the t-product. A function $f : U \to \mathbb{K}$ is called $\mathbb{K}$-*smooth* if its transform-domain components are C^β-smooth, i.e.,

$$[Mf]_k \in C^\beta([M(U)]_k) \quad \text{for } \mathbb{K} = \mathbb{K}_c, \qquad [Mf](t) \in C^\beta([M(U)](t)) \quad \text{for } \mathbb{K} = \mathbb{K}_\infty.$$

This notion extends classical smoothness to the structured t-scalar algebra by enforcing frequency-wise regularity consistent with the t-product[5].

Definition 3 (Structure Sheaf)**.** Let $\mathcal{M}$ be a C^β-smooth t-manifold over $\mathbb{K}$. The structure sheaf $\mathcal{O}_\mathcal{M}$ assigns to each open set $U \subset \mathcal{M}$ the set of all $\mathbb{K}$-smooth functions $f : U \to \mathbb{K}$:

$$\mathcal{O}_\mathcal{M}(U) := \{\, f : U \to \mathbb{K} \mid f \text{ is } \mathbb{K}\text{-smooth} \,\}.$$

This generalizes the classical ring of smooth real-valued functions to structured scalars in $\mathbb{K}$ and serves as the algebraic base for all geometric constructions. Elements of $\mathcal{O}_\mathcal{M}(U)$ can be viewed as generalized smooth scalar fields that take values in the t-product algebra $\mathbb{K}$, respecting smoothness in the transform domain. Having defined the structure sheaf $\mathcal{O}_\mathcal{M}$, we now introduce the tangent space, which captures infinitesimal variations of $\mathbb{K}$-smooth functions on t-manifolds via $\mathbb{K}$-linear derivations.

Definition 4 (Tangent Space)**.** The tangent space $\mathcal{T}_\mathcal{M}(U)$ over $U \subset \mathcal{M}$ is the space of $\mathbb{K}$-linear derivations acting on $\mathcal{O}_\mathcal{M}(U)$:

$$\mathcal{T}_\mathcal{M}(U) := \left\{ D : \mathcal{O}_\mathcal{M}(U) \to \mathcal{O}_\mathcal{M}(U) \;\middle|\; \begin{array}{l} D(f * g) = D(f) * g + f * D(g), \\ D(a * f + b * g) = a * D(f) + b * D(g), \quad \forall a, b \in \mathbb{K} \end{array} \right\}.$$

Each D generalizes a directional derivative in the $\mathbb{K}$-valued setting and plays the role of a vector field.

[5] The definition of $\mathbb{K}$-smooth functions is motivated by the notions of *tubal functions* [49] and *t-functions* [43]. Throughout this work, all $\mathbb{K}$-smooth functions and related geometric objects are assumed to satisfy frequency-wise smoothness in the transform domain. This assumption is considerably stronger than that of general vector-valued smooth functions, which may allow inter-frequency dependencies. While it simplifies both definitions and theoretical analysis, it also limits the expressiveness of the framework for modeling more complex cross-frequency behaviors. Relaxing the frequency-wise smoothness assumption therefore represents a promising direction for future research.

Definition 5 (Differential 1-Forms)**.** The module of 1-forms $\Omega^1_{\mathcal{M}}(U)$ is defined as the Kähler differential module over $\mathcal{O}_{\mathcal{M}}(U)$, generated by formal symbols df subject to the Leibniz rule:

$$d(f * g) = f * dg + g * df.$$

It satisfies $df(X) := X(f)$ for all $X \in \mathcal{T}_{\mathcal{M}}(U)$ and $f \in \mathcal{O}_{\mathcal{M}}(U)$, where $\mathcal{T}_{\mathcal{M}}(U)$ denotes the module of $\mathbb{K}$-linear derivations on $\mathcal{O}_{\mathcal{M}}(U)$.

These definitions establish the algebraic foundation for Riemannian geometry on t-manifolds, including Riemannian metric, gradients, Laplacians, and geodesics, which we develop next.

Riemannian geometry on t-manifolds. We equip each tangent space with a Riemannian metric that respects the t-scalar algebra $\mathbb{K} \in \{\mathbb{K}_c, \mathbb{K}_\infty\}$ and transform-domain smoothness, enabling computation of distances, angles, and gradients.

Definition 6 (Riemannian Metric on t-Manifolds)**.** A *Riemannian metric* on a smooth t-manifold $\mathcal{M}$ is a symmetric, $\mathbb{K}$-positive-definite, and $\mathbb{K}$-bilinear form

$$g : \mathcal{T}_{\mathcal{M}}(U) \times \mathcal{T}_{\mathcal{M}}(U) \to \mathcal{O}_{\mathcal{M}}(U)$$

defined on each open set $U \subset \mathcal{M}$, satisfying the following conditions:

(1) Symmetry: For all $X, Y \in \mathcal{T}_{\mathcal{M}}(U)$ and every $x \in U$, the metric satisfies $g_x(X, Y) = g_x(Y, X)$, where $g_x(X, Y) := g(X, Y)(x) \in \mathbb{K}$ denotes the value of the $\mathbb{K}$-valued function $g(X, Y)$ at x.

(2) $\mathbb{K}$-positive-definite: For every $x \in U$ and nonzero $X \in \mathcal{T}_{\mathcal{M}}(U)$, the element $g_x(X, X) \in \mathbb{K}$ is pointwise positive in the transform domain. That is, $[M g_x(X, X)]_k > 0$ for all k when $\mathbb{K} = \mathbb{K}_c$, and $[M g_x(X, X)](t) > 0$ for all t when $\mathbb{K} = \mathbb{K}_\infty$.

(3) Smoothness: For all $X, Y \in \mathcal{T}_{\mathcal{M}}(U)$, $g(X, Y) \in \mathcal{O}_{\mathcal{M}}(U)$, i.e., it is $\mathbb{K}$-smooth.

Intuitively, a Riemannian metric on a t-manifold assigns an inner product between vector fields, but with values in the t-scalar ring $\mathbb{K}$. In the transform domain, this corresponds to a family of classical real-valued inner products computed slice-wise—one for each frequency index k (in the discrete case) or each parameter t (in the continuous case)—thus preserving compatibility with the algebraic structure of the t-product [32, 33].

The Riemannian metric enables the definition of gradients, which are essential for optimization tasks on t-manifolds. Since functions on $\mathcal{M}$ are $\mathbb{K}$-valued, the gradient must be defined in a way that respects both the algebraic structure of the t-product and the underlying transform.

Definition 7 (Gradient)**.** Let $\mathcal{M}$ be a t-manifold with Riemannian metric g. The gradient ∇f of a function $f \in \mathcal{O}_{\mathcal{M}}(U)$ is the vector field in $\mathcal{T}_{\mathcal{M}}(U)$ satisfying $g(\nabla f, X) = df(X)$ for all $X \in \mathcal{T}_{\mathcal{M}}(U)$.

While defined abstractly, the gradient operates slice-wise in the transform domain, preserving the structure of t-scalars and enabling differentiable learning in frequency-aligned tensor spaces.

Definition 8 (Divergence and Laplacian)**.** Let $X \in \mathcal{T}_{\mathcal{M}}(U)$. The divergence $\operatorname{div}(X)$ is defined via the Lie derivative of the volume form induced by g. The Laplacian of a function f is defined as $\Delta f := \operatorname{div}(\nabla f)$.

The Laplacian governs harmonicity and diffusion on t-manifolds and reduces to classical Laplace operators slice-by-slice in the transformed domain. This structure enables geometric regularization and smoothing in tensor spaces. In particular, it provides a principled generalization of spectral methods used in t-SVD-based graph models [13, 51], where our construction now endows such models with intrinsic manifold-aware Laplacians.

Definition 9 (Levi-Civita Connection and Geodesics)**.** A connection ∇ on $\mathcal{T}_{\mathcal{M}}$ is called the Levi-Civita connection if it is torsion-free and metric-compatible: $\nabla_X Y - \nabla_Y X = [X, Y]$ and $X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$. A curve $\gamma : I \to \mathcal{M}$ is a geodesic if it satisfies $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$.

This generalizes shortest-path and constant-velocity flows to the t-manifold setting. In practice, geodesics respect the transform structure and can be computed slice-wise, providing a bridge to structure-aware modeling in tensor dynamics.

These constructions establish a differential geometric framework for t-manifolds, addressing **Q1** by unifying discrete and continuous tensor data with transform-consistent metrics, operators, and flows which are crucial for *extending traditional manifold learning techniques [46] to the t-scalar setting*. This foundation also enables learning theory on t-manifold (**Q2**) investigated in the next section.

3 Learning Theory on t-Manifolds: Testing, Fitting, and Function Learning

Leveraging the t-manifold geometry framework, we tackle **Q2** by testing and fitting a low-dimensional t-manifold, then modeling its functions with tensor neural networks.

3.1 Theory of t-Manifold Hypothesis Testing and Fitting

We begin our study of **Q2** by examining the geometric structure underlying tensor data. Motivated by the manifold hypothesis [18], we ask: *can t-vector data in $\mathbb{K}_c^d$ be well approximated by a low-dimensional smooth t-manifold?*

When tensor-valued data concentrate around a low-dimensional t-manifold $\mathcal{M}$, learning can be performed in a reduced, structure-aligned space, which enhances generalization and interpretability. To model such structure, we exploit the algebraic property of $\mathbb{K}_c = (\mathbb{R}^{1\times 1\times c}, +, *)$, where the t-product is defined through an orthogonal transform $M \in \mathbb{R}^{c\times c}$ that decouples tensor operations into frequency-wise matrix multiplications. This yields a natural frequency-domain representation: a t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$ is characterized by its frequency slices $\mathcal{M}_k := \{[Mx]_k : x \in \mathcal{M}\} \subset \mathbb{R}^d$, which jointly describe the global tensor geometry.

Specifically, we extend the manifold hypothesis framework [18] to the t-product setting. Given samples $\{x_i\}_{i=1}^n \subset \mathbb{K}_c^d$, we test whether, for each frequency index k, the transformed slice $\{[Mx_i]_k\}_{i=1}^n$ lies near a p-dimensional manifold $\mathcal{M}_k$. The global t-manifold candidate is then defined as the inverse transform of the slice manifold collection $\{\mathcal{M}_k\}_{k=1}^c$. To quantify data concentration, we aggregate the per-slice deviations $\sum_{k=1}^c \mathrm{dist}([Mx_i]_k, \mathcal{M}_k)^2$ and test whether their average remains below a given resolution ϵ.

To formally characterize the complexity of a t-manifold $\mathcal{M}$, we define its volume and reach via its transform-domain components:

Definition 10 (Spectral Volume and Reach)**.** Let $\mathcal{M} \subset \mathbb{K}_c^d$ be a p-dimensional t-manifold, and let $\mathcal{M}_k := \{[Mx]_k : x \in \mathcal{M}\} \subset \mathbb{R}^d$ denote its k-th frequency slice. We define the *spectral volume* and *spectral reach* of $\mathcal{M}$ as $\mathrm{Vol}_{\mathrm{spec}} := \sum_{k=1}^c \mathcal{H}^p_{\mathbb{R}^d}(\mathcal{M}_k)$ and $\mathrm{reach}_{\mathrm{spec}} := \min_{1\leqslant k\leqslant c} \mathrm{reach}_{\mathbb{R}^d}(\mathcal{M}_k)$, where $\mathcal{H}^p_{\mathbb{R}^d}$ denotes the p-dimensional Hausdorff measure in $\mathbb{R}^d$, and $\mathrm{reach}_{\mathbb{R}^d}$ is the reach measured in the Euclidean metric.

Theorem 1 (t-Manifold Hypothesis Testing)**.** *Let $\{x_i\}_{i=1}^n \subset \mathbb{B}_{\mathbb{K}_c^d}$ be i.i.d. samples normalized to the unit ball.*[6] *Fix intrinsic dimension p, upper bound V on the spectral volume, lower bound τ on the spectral reach, resolution $\epsilon > 0$, and confidence level $\delta \in (0, 1)$. Then there exists a statistical test that, for sufficiently large n, distinguishes with probability at least $1 - \delta$ between the following two situations:*

*(I) (*Near case*) There exists $\mathcal{M} \in \mathcal{G}(p, CV, \tau/C)$ such that $\frac{1}{n}\sum_{i=1}^n \mathrm{dist}^2(x_i, \mathcal{M}) \leqslant C\epsilon$.*

*(II) (*Far case*) For all $\mathcal{M} \in \mathcal{G}(p, V/C, C\tau)$, $\frac{1}{n}\sum_{i=1}^n \mathrm{dist}^2(x_i, \mathcal{M}) > \epsilon/C$.*

Here, $\mathcal{G}(p, V, \tau)$ denotes the class of C^2 t-manifolds with spectral volume at most V and spectral reach at least τ, and C is a positive constant depending only on p and c.

This result suggests that low-dimensional t-manifold structures can, in principle, be detected from high-dimensional tensor data, even when the underlying geometry is implicit and embedded in a structured transform-based representation.

While Theorem 1 confirms the presence of t-manifold structure, addressing **Q2** further requires reconstructing the manifold itself. We now consider *the fitting problem*: *given noisy samples $\{x_i\}_{i=1}^n \subset \mathbb{K}_c^d$ near a p-dimensional smooth t-manifold, can we recover a smooth estimator that approximates it up to geometric accuracy*?

Theorem 2 (t-Manifold Fitting)**.** *Let $x_i = z_i + \xi_i \in \mathbb{K}_c^d$ for $i = 1, \ldots, n$, where $z_i \sim \mathrm{Unif}(\mathcal{M})$ is sampled from a t-manifold $\mathcal{M} \in \mathcal{G}(p, V, \tau)$, and $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{cd})$ represents additive Gaussian noise. Assume the manifold dimension p and noise level σ are known. If $n = \tilde{\mathcal{O}}(\sigma^{-(p+3)})$ and σ*

[6] Rescaling so that $\|x_i\|_{\mathbb{K}_c^d} \leqslant 1$ does not affect the generality of the result, since both spectral reach and volume scale homogeneously under dilation.

is sufficiently small, then with high probability, the estimator $\widehat{\mathcal{M}}$ is a p-dimensional C^2 t-manifold embedded in $\mathbb{K}_c^d$ satisfying the Hausdorff distance bound:

$$d_H(\mathcal{M}, \widehat{\mathcal{M}}) \leqslant C\sigma^2 \log(1/\sigma),$$

where the distance is measured under the $\mathbb{K}_c^d$ norm $\|x\|_{\mathbb{K}_c^d} := \left(\sum_{k=1}^c \|[Mx]_k\|_2^2\right)^{1/2}$, and the constant C depends on c, p, V, τ, and the unitary transform M.

The theorem establishes the pointwise and global proximity of $\widehat{\mathcal{M}}$ to the target t-manifold while preserving its C^2 regularity. This underscores the consistency of the proposed contraction-based estimator within the $\mathbb{K}_c^d$ geometric setting.

3.2 Function Learning Theory on t-Manifolds

To further address **Q2**, we consider the problem of learning functions defined on a p-dimensional t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$, which may be a suitable model in structured prediction tasks such as image or video analysis. The goal is to approximate an unknown $\mathbb{K}_c$-smooth function $f_0 : \mathcal{M} \to \mathbb{K}_c$ that maps high-dimensional t-vector inputs to structured t-scalar outputs, such as pixel intensities or compressed features. While the low-dimensional geometry of $\mathcal{M}$ enables efficient representation, the algebraic complexity of the t-scalar ring $\mathbb{K}_c$, defined via transform-based multiplication, poses unique challenges. In particular, any approximation architecture must respect both the non-Euclidean geometry of $\mathcal{M}$ and the algebraic structure of $\mathbb{K}_c$.

To address this, we adopt tensor neural networks (TNNs) specifically designed to operate over t-product spaces [44, 61, 48]. A TNN approximator is constructed as a composition of L layers:

$$h_l = \hat{\sigma}(W_l * h_{l-1}), \quad l = 1, \ldots, L, \tag{1}$$

where $h_0 = x_i \in \mathbb{K}_c^d$ is the input tensor, $W_l \in \mathbb{R}^{m_l \times m_{l-1} \times c}$ are learnable weight tensors, and the activation function is defined as $\hat{\sigma}(x) = \sigma(x) \times_3 M$, where σ denotes the elementwise ReLU function. This transform-domain activation design follows [44], and $M \in \mathbb{R}^{c \times c}$ is a fixed orthogonal transform, such as the DCT.

The network outputs a function $\hat{f}(x_i; \theta) := h_L \in \mathbb{K}_c$, where $\theta = \{W_1, \ldots, W_L\}$ denotes all model parameters. Training is performed via empirical risk minimization under the observation model:

$$y_i = f_0(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2}$$

where $y_i \in \mathbb{K}_c$ is the response and ϵ_i denotes zero-mean Gaussian noise with covariance $\sigma^2 I_c$. The empirical objective is defined directly in the transform domain:

$$\hat{f}_n \in \arg\min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \|y_i - f(x_i)\|_{\mathbb{K}_c}^2 := \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c \left|[M(y_i - f(x_i))]_k\right|^2. \tag{3}$$

This formulation preserves transform consistency and ensures compatibility with the t-product structure of $\mathbb{K}_c$. It also reveals the core challenge of t-manifold learning: designing neural architectures capable of approximating $\mathbb{K}_c$-valued functions over domains that are both geometrically curved and algebraically structured.

We introduce regularity assumptions under which TNNs admit provable generalization guarantees.

Assumption 1 (Modeling Assumptions)**.** The following conditions hold:

(A1) *Data distribution regularity:* The data points $\{x_i\} \subset \mathbb{K}_c^d$ lie within a compact p-dimensional C^β t-manifold $\mathcal{M}$ with spectral reach at least τ and spectral volume at most V. The sampling distribution ν is supported on a compact subset of $\mathcal{M}$ and is absolutely continuous with respect to the spectral volume measure.

(A2) *Target function smoothness:* The function $f_0 : \mathcal{M} \to \mathbb{K}_c$ is C^β-$\mathbb{K}_c$-smooth with bounded norm:

$$\|f_0\|_{C^\beta(\mathcal{M})} := \sup_{\alpha, k, \|\alpha\|_1 \leqslant \beta} \sup_{\mathbf{x} \in \varphi_\alpha(U_\alpha)} |\partial^\alpha [M(f_0 \circ \varphi_\alpha^{-1})]_k(\mathbf{x})| \leqslant B_0.$$

(A3) *Model class constraint:* The hypothesis class $\mathcal{F}_n$ consists of TNNs with width N_0, depth L_0, total parameter count $\mathcal{S}$, and output norm bound $\mathcal{B}$, using the t-product defined via a fixed orthogonal transform M.

Remark 1. Assumption 1 reflects natural and interpretable conditions adapted to the spectral geometry of t-manifolds: Specifically, *(A1)* extends the classical manifold hypothesis to the t-product setting via spectral volume and reach [29]; *(A2)* requires frequency-slice smoothness of f_0, a t-product variant of a standard assumption in geometric deep learning [35]; *(A3)* reflects practical TNN design and enforces algebraic compatibility with $\mathbb{K}_c$ [44, 61].

We are now ready to state the convergence guarantee for TNNs trained via empirical risk minimization on t-manifolds.

Theorem 3 (TNN Approximation on t-Manifolds)**.** *Define* $p_{\text{eff}} p_{\text{eff}} = O(p \log(cd))$ *as the effective dimension determined by the intrinsic complexity of* $\mathcal{M}$*. Under Assumption 1, there exists a TNN class* $\mathcal{F}_n$ *such that the empirical risk minimizer* $\hat{f}_n$ *of Problem (3) satisfies*

$$\mathbb{E}\|\hat{f}_n - f_0\|^2_{L^2(\nu)} \leqslant C n^{-\frac{2\beta}{p_{\text{eff}}+2\beta}},$$

where C *depends on* $(p, B_0, V, \tau, \sigma, c, d, \log n)$*. Here,* $\mathbb{E}\|\hat{f}_n - f_0\|^2_{L^2(\nu)}$ *represents the expected squared* L^2*-error between the TNN estimator* $\hat{f}_n$ *and the true target function* f_0*, measured with respect to the sampling distribution* ν *on* $\mathcal{M}$*.*

This result shows that TNNs can approximate smooth functions on t-manifolds with rates determined by the effective dimension $\mathcal{O}(p \log(cd))$. The sample-dependent term benefits from the low effective complexity of the t-manifold, achieving a near-optimal nonparametric regression rate under strict manifold support. The bound scales polynomially with ambient and spectral parameters, aligning with classical manifold approximation theory [29, 35] and extending it to the t-manifold settings.

4 Modeling Implications of t-Product Geometry

This section examines how the framework of t-product geometry can *inspire new modeling perspectives*, complementing the theoretical results developed in Sections 2 and 3. Rather than validating the theory through experiments, we focus on clarifying how the algebraic structures of t-scalars, t-modules, and t-manifolds translate into modeling constraints and design principles for high-dimensional tensor data, such as how linearity, locality, and curvature are represented in the transform domain.

At the foundational level, the t-scalar serves as a basic modeling unit, such as a row or column of an image [83], while preserving its internal spectral structure. Building upon this, the t-module provides a flat t-linear space formed by linear combinations of t-scalars. At a more general level, t-product geometry allows for nonlinear structures, such as curvature or twisting, to arise when linear t-modules are combined, constrained, or glued together in nontrivial ways. From this viewpoint, the central modeling perspective of t-product geometry is to *understand how structured constraints on t-scalar representations give rise to effective geometric structure*.

Building on this understanding, we explore how t-product geometry can inform model construction. Section F.1 introduces the Bidirectional Tensor Representation (BTR) formulation, which incorporates dual t-module constraints for structured learning tasks such as clustering and tensor recovery. Preliminary evaluations across several data modalities, including images, videos, hyperspectral and multispectral images, point clouds, and thermal sequences, suggest that BTR offers a coherent and flexible modeling framework. Broader extensions and discussions are provided in Section F.2.

Example: Bidirectional Tensor Representation (BTR). Images naturally exhibit row–column symmetry, which aligns with the dual-module structure implied by the t-product: a two-dimensional image can be regarded as a t-vector in either $\mathbb{K}_w^h$ (row-wise) or $\mathbb{K}_h^w$ (column-wise). The BTR formulation leverages this bidirectional structure by applying low-rank regularization in both modules through tensor nuclear norm surrogates, thereby promoting coherence from both perspectives.

Although $\mathbb{K}_w^h$ and $\mathbb{K}_h^w$ are each flat (linear) modules, enforcing low-rankness jointly in both induces a coupling that geometrically *twists* the representation space. From a geometric viewpoint, this coupling gives rise to an effective, constraint-induced nonlinearity in the representation space, which is empirically associated with improved generalization. The objective formulation and optimization details are presented in Appendix F.1.

We evaluate BTR on image clustering and video denoising tasks. As shown in Table 1, BTR achieves consistent gains across clustering metrics (ACC, NMI, PUR) and yields the highest PSNR values in YUV video denoising under 20% noise (Figure F.2). More experiments on Poisson tensor completion are provided in Appendix F.1.3. These results illustrate how incorporating geometric principles from the t-product perspective can enhance structured tensor modeling.

Table 1: Clustering performance comparison on five benchmark datasets using accuracy (ACC), normalized mutual information (NMI), and purity (PUR).

Dataset	Metric	R-TPCA		OR-TPCA	R-TLRR		OR-TLRR		BTR (Proposed)	
		DFT [41]	**DCT** [40]	[82]	**DFT** [83]	**DCT** [73]	**DFT** [68]	**DCT** [68]	**DFT**	**DCT**
FRDUE	ACC	0.7613	0.7777	0.7644	0.8429	0.8400	0.8358	0.7386	0.8591	**0.8594**
	NMI	0.9093	0.9140	0.9127	0.9511	0.9510	0.9470	0.9045	**0.9578**	0.9567
	PUR	0.7955	0.8077	0.7990	0.8760	0.8718	0.8643	0.7776	**0.8901**	0.8890
FRDUE-100	ACC	0.7906	0.7911	0.7974	0.8606	0.8602	0.8657	0.7616	**0.8860**	0.8769
	NMI	0.9133	0.9136	0.9221	0.9529	0.9526	0.9564	0.9055	**0.9643**	0.9605
	PUR	0.8197	0.8200	0.8271	0.8882	0.8881	0.8926	0.7956	**0.9128**	0.9032
Olivetti	ACC	0.3970	0.3703	0.3965	0.5230	0.6452	0.6162	0.5535	0.5645	**0.6670**
	NMI	0.5990	0.5809	0.5987	0.6920	0.7905	0.7732	0.7395	0.7290	**0.8080**
	PUR	0.4242	0.3983	0.4200	0.5483	0.6755	0.6492	0.5880	0.5960	**0.7005**
PIE-10	ACC	0.4276	0.4268	0.5401	0.5897	0.5831	0.4594	0.1975	0.6000	**0.6110**
	NMI	0.6674	0.6621	0.7361	0.7562	0.7618	0.7049	0.5150	0.7697	**0.7778**
	PUR	0.4469	0.4462	0.5593	0.6059	0.5999	0.4803	0.2050	0.6187	**0.6325**
USPS1000	ACC	0.3548	0.3537	0.3369	0.4257	0.4088	0.5191	0.4499	**0.5799**	0.5339
	NMI	0.3066	0.2986	0.2828	0.3834	0.3915	0.5104	0.4490	**0.5860**	0.5390
	PUR	0.4470	0.4542	0.4425	0.5264	0.5076	0.6330	0.5905	**0.6862**	0.6662

5 Concluding Remarks

We introduce a general framework of t-product geometry that extends differential geometry to the t-product algebra for tensor learning. Through a sheaf-theoretic formulation, we define metrics, gradients, Laplacians, and geodesics over both discrete and continuous t-scalars, unifying transform-domain representations within a coherent geometric structure. To our knowledge, this is the first systematic development of differential geometry on t-scalars, enabling structured modeling of tensor data. We also present a theoretical study of learning on t-manifolds, encompassing hypothesis testing, manifold fitting, and function learning. The potential applicability of this framework is illustrated through examples in image clustering and video denoising. These results offer a principled foundation for geometry-aware tensor learning.

Limitations and future work. This work focuses on the theoretical foundations of t-product geometry, emphasizing conceptual definitions, assumptions, and provable guarantees rather than empirical evaluations. ***The authors believe that developing a mature and comprehensive theory in this direction is a highly nontrivial task***. The framework presented here is *preliminary* and *necessarily incomplete*, yet it aims to shed light on what a general theory of t-product geometry might entail and to provide a foundation for continued theoretical development in this direction.

This work has limitations in scope, idealized assumptions, and algorithmic development:

- **Scope of study.** This paper focuses primarily on the theoretical formulation of t-product geometry rather than empirical validation. A natural next step lies in extending the framework to richer modeling paradigms, such as *generative modeling, geometric graph networks, temporal dynamics, manifold optimization, manifold learning, and federated tensor learning*, where geometric consistency across spectral modes could offer new algorithmic principles and inductive biases.
- **Idealized assumptions.** For theoretical clarity, several idealized assumptions were made: (1) transform-domain smoothness, (2) the existence of a unit element in the t-scalar ring, and (3) frequency-wise independence among spectral components. These assumptions make analysis tractable but restrict realism. In particular, relaxing the independence assumption to allow structured cross-frequency coupling may reveal richer geometric behaviors and better reflect the complexity of real-world tensor data.
- **Algorithmic extensions.** Beyond theory, the computational side of t-product geometry remains largely unexplored. Developing numerically stable and scalable algorithms that leverage the proposed geometric structures, for instance in *optimization over t-manifolds, spectral regularization, or tensor-valued neural architectures*, poses both challenges and opportunities.

Acknowledgments and Disclosure of Funding

The authors sincerely thank the Area Chair and the four anonymous reviewers for their detailed and constructive feedback. Their suggestions have greatly enhanced the quality and clarity of this paper. This work was supported in part by the National Natural Science Foundation of China under Grant Numbers 62203124, 62562065; JSPS KAKENHI Grant Numbers JP25K21283, JP24K20849, JP23K28109, JP24K03005, JP25K21288; JSPS Bilateral Program Number JPJSBP120257420; and RIKEN Incentive Research Project 100847-202301062011. Yuning Qiu was supported by the RIKEN Special Postdoctoral Researcher Program.

References

[1] N. Aigerman, K. Gupta, V. G. Kim, S. Chaudhuri, J. Saito, and T. Groueix. Neural jacobian fields: learning intrinsic mappings of arbitrary meshes. *ACM Transactions on Graphics (TOG)*, 41(4):1–17, 2022.

[2] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, Cambridge, 1999.

[3] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.

[4] P. L. Bartlett, N. Harvey, C. Liaw, and A. Mehrabian. Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.*, 20:Paper No. 63, 17, 2019.

[5] M. Belkin and P. Niyogi. Semi-supervised learning on riemannian manifolds. *Machine learning*, 56(1):209–239, 2004.

[6] K. Braman. Third-order tensors as linear operators on a space of matrices. *Linear Algebra and its Applications*, 433(7):1241–1253, 2010.

[7] G. E. Bredon. *Sheaf Theory*, volume 170 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 1997.

[8] Y. Cao and Y. Xie. Poisson matrix recovery and completion. *IEEE Transactions on Signal Processing*, 64(6):1609–1620, 2015.

[9] M. Chen, H. Jiang, W. Liao, and T. Zhao. Nonparametric regression on low-dimensional manifolds using deep relu networks. *arXiv:1908.01842*, 2019.

[10] M. Chen, H. Jiang, and T. Zhao. Efficient approximation of deep relu networks for functions on low dimensional manifolds. *Advances in Neural Information Processing Systems*, 2019.

[11] H. Chung, B. Sim, D. Ryu, and J. C. Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.

[12] J. W. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer vision and image understanding*, 106(2-3):162–182, 2007.

[13] L. Deng, X.-Y. Liu, H. Zheng, X. Feng, and Y. Chen. Graph spectral regularized tensor completion for traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):10996–11010, 2021.

[14] S. Esposito, Q. Xu, K. Kania, C. Hewitt, O. Mariotti, L. Petikam, J. Valentin, A. Onken, and O. Mac Aodha. Geogen: Geometry-aware generative modeling via signed distance functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7479–7488, 2024.

[15] C. Fefferman. Whitney's extension problem for c^m. *Annals of Mathematics.*, 164(1):313–359, 2006.

[16] C. Fefferman, S. Ivanov, Y. Kurylev, M. Lassas, and H. Narayanan. Fitting a putative manifold to noisy data. In *Conference On Learning Theory*, pages 688–720. PMLR, 2018.

[17] C. Fefferman, S. Ivanov, M. Lassas, and H. Narayanan. Fitting a manifold of large reach to noisy data. *Journal of Topology and Analysis*, 17(02):315–396, 2025.

[18] C. Fefferman, S. Mitter, and H. Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[19] K. Gilman, D. A. Tarzanagh, and L. Balzano. Grassmannian optimization for online tensor completion and tracking with the t-SVD. *IEEE transactions on signal processing*, 70:2152–2167, 2022.

[20] A. Grothendieck and J. Dieudonné. Eléments de géométrie algébrique. 1964.

[21] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer-Verlag, New York, 2002.

[22] R. Hartshorne. *Algebraic Geometry*, volume 52 of *Graduate Texts in Mathematics*. Springer, 1977.

[23] 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.

[24] Z. Huang, X. Li, Y. Ye, and M. K. Ng. Mr-gcn: Multi-relational graph convolutional networks based on generalized tensor product. In *IJCAI*, volume 20, pages 1258–1264, 2020.

[25] B. Hui, Z. Song, H. Fan, P. Zhong, W. Hu, X. Zhang, J. Ling, H. Su, W. Jin, Y. Zhang, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. *China Sci. Data*, 5(3):291–302, 2020.

[26] A. J. Izenman. Introduction to manifold learning. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(5):439–446, 2012.

[27] J. Ji and S. Feng. Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19343–19352, 2023.

[28] Q. Jiang and M. Ng. Robust low-tubal-rank tensor completion via convex optimization. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 2649–2655. AAAI Press, 2019.

[29] Y. Jiao, G. Shen, Y. Lin, and J. Huang. Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors. *The Annals of Statistics*, 51(2):691–716, 2023.

[30] E. Kernfeld, S. Aeron, and M. Kilmer. Clustering multi-way data: A novel algebraic approach. *arXiv preprint arXiv:1412.7056*, 2014.

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

[32] 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.

[33] M. E. Kilmer, L. Horesh, H. Avron, and E. Newman. Tensor-tensor algebra for optimal representation and compression of multiway data. *Proceedings of the National Academy of Sciences*, 118(28):e2015851118, 2021.

[34] W. Kühnel. *Differential geometry*, volume 77. American Mathematical Soc., 2015.

[35] D. Labate and J. Shi. Optimal approximation of smooth functions on riemannian manifolds using deep relu neural networks. *Available at SSRN 5117796*.

[36] J. M. Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

[37] X. Li, M. K. Ng, G. Xu, and A. Yip. Multi-relational graph convolutional networks: Generalization guarantees and experiments. *Neural Networks*, 161:343–358, 2023.

[38] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE TPAMI*, 35(1):208–220, 2013.

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

[40] C. Lu. Transforms based tensor robust pca: Corrupted low-rank tensors recovery via convex optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1145–1152, 2021.

[41] C. Lu, J. Feng, W. Liu, Z. Lin, S. Yan, et al. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE TPAMI*, 2019.

[42] 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.

[43] K. Lund. The tensor t-function: A definition for functions of third-order tensors. *Numerical Linear Algebra with Applications*, 27(3):e2288, 2020.

[44] O. A. Malik, S. Ubaru, L. Horesh, M. E. Kilmer, and H. Avron. Dynamic graph convolutional networks using the tensor m-product. In *Proceedings of the 2021 SIAM international conference on data mining (SDM)*, pages 729–737. SIAM, 2021.

[45] X.-P. Mao, Y. Wang, and Y.-N. Yang. Computation over t-product based tensor stiefel manifold: A preliminary study. *Journal of the Operations Research Society of China*, pages 1–49, 2024.

[46] M. Meilă and H. Zhang. Manifold learning: What, how, and why. *Annual Review of Statistics and Its Application*, 11(1):393–417, 2024.

[47] U. Mor and H. Avron. Quasitubal tensor algebra over separable hilbert spaces. *arXiv preprint arXiv:2504.16231*, 2025.

[48] E. Newman, L. Horesh, H. Avron, and M. Kilmer. Stable tensor neural networks for rapid deep learning. *arXiv preprint arXiv:1811.06569*, 2018.

[49] E. Newman, L. Horesh, H. Avron, and M. E. Kilmer. Stable tensor neural networks for efficient deep learning. *Frontiers in Big Data*, 7:1363978, 2024.

[50] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

[51] K. Pena-Pena, D. L. Lau, and G. R. Arce. T-HGSP: Hypergraph signal processing using t-product tensor decompositions. *IEEE Transactions on Signal and Information Processing over Networks*, 9:329–345, 2023.

[52] F. Qian, Z. Liu, Y. Wang, Y. Zhou, and G. Hu. Ground truth-free 3-d seismic random noise attenuation via deep tensor convolutional neural networks in the time-frequency domain. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.

[53] H. Qiu, Y. Wang, S. Tang, D. Meng, and Q. Yao. Fast and provable nonconvex tensor RPCA. In *International Conference on Machine Learning*, pages 18211–18249. PMLR, 2022.

[54] 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.

[55] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

[56] T. Sakai. *Riemannian geometry*, volume 149. American Mathematical Soc., 1996.

[57] G. Sardanashvily. *Lectures on differential geometry of modules and rings: Application to Quantum Theory*. LAP LAMBERT Academic Publishing, 2012.

[58] J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function (with discussion). *Ann. Statist.*, 48(4):1875–1897, 2020.

[59] G. Song, M. K. Ng, and X. Zhang. Robust tensor completion using transformed tensor singular value decomposition. *Numerical Linear Algebra With Applications*, 27(3):e2299, 2020.

[60] R. Tang and Y. Yang. Adaptivity of diffusion models to manifold structures. In *International Conference on Artificial Intelligence and Statistics*, pages 1648–1656. PMLR, 2024.

[61] 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. In *Advances in Neural Information Processing Systems*, 2023.

[62] A. Wang, Y. Qiu, M. Bai, Z. Jin, G. Zhou, and Q. Zhao. Generalized tensor decomposition for understanding multi-output regression under combinatorial shifts. In *Advances in Neural Information Processing Systems*, 2024.

[63] A. Wang, Y. Qiu, H. Huang, Z. Jin, G. Zhou, and Q. Zhao. Refining dual spectral sparsity in transformed tensor singular values, 2025.

[64] A. Wang, Y. Qiu, Z. Jin, G. Zhou, and Q. Zhao. Low-rank tensor transitions (LoRT) for transferable tensor regression. In *International Conference on Machine Learning*, volume 267, 2025.

[65] 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.

[66] C. Wang, X.-L. Zhao, Y.-B. Zheng, B.-Z. Li, and M. K. Ng. Functional tensor singular value decomposition. *SIAM Journal on Scientific Computing*, 47(4):A2180–A2204, 2025.

[67] J. Wang and X. Zhao. Functional transform-based low-rank tensor factorization for multi-dimensional data recovery. In *European Conference on Computer Vision*, pages 39–56. Springer, 2024.

[68] T. Wu. Robust data clustering with outliers via transformed tensor low-rank representation. In *International Conference on Artificial Intelligence and Statistics*, pages 1756–1764. PMLR, 2024.

[69] T. Wu and W. U. Bajwa. A low tensor-rank representation approach for clustering of imaging data. *IEEE Signal Processing Letters*, 25(8):1196–1200, 2018.

[70] Z. Wu, L. Shu, Z. Xu, Y. Chang, C. Chen, and Z. Zheng. Robust tensor graph convolutional networks via t-SVD based graph augmentation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2090–2099, 2022.

[71] 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.

[72] 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.

[73] J.-H. Yang, C. Chen, H.-N. Dai, M. Ding, Z.-B. Wu, and Z. Zheng. Robust corrupted data recovery and clustering via generalized transformed tensor low-rank representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[74] Z. Yao, J. Su, B. Li, and S.-T. Yau. Manifold fitting. *arXiv preprint*, 2023. arXiv:2304.07680.

[75] Z. Yao, J. Su, and S.-T. Yau. Manifold fitting with cyclegan. *Proceedings of the National Academy of Sciences*, 121(5):e2311436121, 2024.

[76] Z. Yao and Y. Xia. Manifold fitting under unbounded noise. *Journal of Machine Learning Research*, 26(45):1–55, 2025.

[77] D. Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017.

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

[79] Z. Zhang and S. Aeron. Exact tensor completion using t-SVD. *IEEE TSP*, 65(6):1511–1526, 2017.

[80] 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.

[81] Z. Zhang, K. Zhang, M. Chen, Y. Takeda, M. Wang, T. Zhao, and Y.-X. Wang. Nonparametric classification on low dimensional manifolds using overparameterized convolutional residual networks. *Advances in Neural Information Processing Systems*, 37:65738–65764, 2024.

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

[83] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan. Tensor low-rank representation for data recovery and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5):1718–1732, 2019.

[84] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

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 three main contributions: (1) the definition of t-manifolds and associated differential geometric structures, (2) a theoretical framework for t-manifold learning including hypothesis testing and function approximation, and (3) a conceptual application illustrating empirical implications. These align with the main technical developments in the paper.

   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 of limitations in the conclusion, acknowledging its primary focus on theoretical development. It highlights several simplifying assumptions made for conceptual clarity, such as transform-domain smoothness and the existence of a unit element. The discussion also notes that complex practical algorithms and broader model classes are beyond the scope of this initial exploration.

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,including the t-manifold hypothesis testing theorem, manifold fitting theorem, and the TNN approximation theorem, the paper clearly states all necessary assumptions and outlines the corresponding results. Full proofs are deferred to the appendix.

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 main experimental results, such as the clustering and denoising case study using BTR, are presented with detailed metrics, dataset names, comparison methods, and tables. The appendix is indicated to provide further details including objective functions and optimization procedures, ensuring reproducibility in relation to the conceptual claims.

   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 made the code publicly available and indicate in the supplemental material that it includes all necessary scripts, data links, and instructions to reproduce the main experimental results such as clustering accuracy and PSNR in denoising, which supports faithful replication.

   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: While the main paper focuses on theoretical contributions, it refers to the appendix for algorithm details. The case study involving the proposed models specifies datasets, metrics, and comparison baselines, and the supplementary material reportedly includes further information on hyperparameters and optimization, satisfying reproducibility requirements.

   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 presents performance metrics (e.g., ACC, NMI, PUR, PSNR) through tables and figures, focusing on conceptual illustration rather than exhaustive benchmarking. As the empirical results aim to demonstrate the potential utility of the proposed framework, statistical uncertainty measures such as error bars or significance tests are not emphasized.

   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 supplemental material includes information about the computational setup, including hardware specifications 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 paper focuses on theoretical and methodological advancements in tensor geometry and learning, uses only public datasets for illustrative experiments, and does not involve sensitive data, human subjects, or deployment risks, thus conforming to 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 is a purely theoretical work focused on foundational geometry and learning theory for tensor data. As such, it does not raise direct societal concerns or impacts that would typically require discussion.

    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: The paper does not involve the release of models or datasets with high risk for misuse. It focuses on theoretical frameworks and illustrative experiments using standard public datasets.

    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: The paper uses publicly available datasets and baseline models, all of which are properly cited in the references. There is no indication of license violations, and terms of use appear to be respected.

    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 the geometry-inspired models and associated code, which is made available with documentation and usage instructions in the supplemental material, supporting reproducibility and clarity for users.

    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: The paper does not involve any crowdsourcing experiments or research with human subjects, so this question is not applicable.

    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: The study does not involve human participants, so there are no associated risks or requirements for IRB approval.

    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: The core method development in this research does not involve LLMs as any important, original, or non-standard components. In accordance with the NeurIPS LLM policy, any use of LLMs was limited to language polishing or code formatting and does not affect the methodology, scientific rigor, or originality of the work.

    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.