

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

This appendix provides the full technical development, theoretical analysis, and supporting case studies for the framework of *t-product geometry* introduced in the main paper. The main text proposes a geometric foundation for tensor learning by extending differential geometry to the algebra induced by the t-product. It addresses two central theoretical questions: **(Q1)** whether one can define a rigorous differential geometric structure over t-scalars that unifies discrete and continuous tensor modeling; and **(Q2)** whether such a structure supports principled analysis of learning tasks on low-dimensional t-manifolds.

To support these goals, the appendix is organized into six sections, reflecting the *algebra–geometry–theory–modeling* progression developed in the main text.

Section A introduces the algebraic preliminaries of the t-product and t-SVD. These constructions define the scalar ring \mathbb{K} and the module operations that form the algebraic basis for our geometric formulation. We also provide a brief overview of related work to situate our approach in the broader tensor learning literature.

Section B provides the complete mathematical foundations for differential geometry on t-manifolds. Building on the t-scalar algebra, we formalize the structure sheaf of \mathbb{K} -valued functions, define the tangent module and Kähler differentials, and extend these constructions to Riemannian metrics, affine connections, and geodesics. This section resolves **Q1** and offers a transform-domain generalization of classical smooth manifolds that respects spectral structure.

Sections C–E develop the theoretical underpinnings of geometry-aware learning on t-manifolds, directly addressing **Q2**. Section **C** establishes a testing framework for validating whether high-dimensional tensor data lies near a t-manifold. Section **D** introduces a principled approach to fitting smooth t-manifolds from noisy data, while Section **E** analyzes the generalization of function approximation over such structures using tensor neural networks. Together, these sections form a coherent theoretical pipeline for t-manifold-based learning.

Section F explores the modeling potential of t-product geometry, showing how constructs like t-modules and bi-module regularization can guide practical tensor modeling. Using the Bidirectional Tensor Representation (BTR) as an example, it demonstrates their effectiveness in tasks like clustering and tensor recovery across diverse data types, including images and hyperspectral data. Broader applications in machine learning are also discussed.

Broader Impacts This work aims to take an initial step toward establishing a geometric framework for tensor learning based on the t-product algebra. It focuses on the theoretical formulation of t-manifolds, differential structures, and learning guarantees, with a preliminary modeling example to illustrate potential use. As the work is primarily conceptual and mathematical, we do not foresee direct positive or negative societal impacts at this stage.

Contents

1	Introduction	1
2	Differential Geometry over the t-Product Algebra	3
2.1	Preliminaries on t-Scalar-Based Representation	3
2.2	Smooth t-Manifolds with Intrinsic Dimension p	4
2.3	Differential Geometry on t-Manifolds	5
3	Learning Theory on t-Manifolds: Testing, Fitting, and Function Learning	7
3.1	Theory of t-Manifold Hypothesis Testing and Fitting	7
3.2	Function Learning Theory on t-Manifolds	8
4	Modeling Implications of t-Product Geometry	9
5	Concluding Remarks	10
A	Preliminaries and Related Works	27
A.1	Preliminaries of t-SVD	27
A.2	Related Work	28
B	The Mathematical Foundations of t-Product Geometry	31
B.1	From Fields to Commutative Rings: The Algebraic Scalar System	31
B.2	From Charts to Modules: The Definition of t-Manifold	31
B.3	From Smooth Functions to Sheaf-Based Derivations: Differential Geometry on t-Manifolds	32
B.3.1	Structure Sheaf, Tangent Derivations, and Kähler Differentials	33
B.3.2	Riemannian Geometry on t-Manifolds	34
B.3.3	Summary and Potential Extensions	36
C	Theoretical Analysis For t-Manifold Hypothesis Testing	37
C.1	General Strategy of the Proof	37
C.2	Geometric and Statistical Lemmas for t-Manifolds	37
C.3	Final Proof of Theorem 1	46
D	Theoretical Analysis For t-Manifold Fitting	48
D.1	Proof Strategy	48
D.2	Synthetic Validation of t-Manifold Fitting	51
E	Theoretical Analysis For Learning on Approximate t-Manifolds	53
E.1	Learning with Approximate t-Manifold Support	54
E.2	Final Proof of the Main Theorem	59
F	Modeling Perspectives and Potential of t-Product Geometry	60

F.1	Example: Bidirectional Tensor Representation (BTR)	60
F.1.1	Motivation: From t-Module to Bidirectional Representation	60
F.1.2	BTR Model for Tensor LRR	61
F.1.3	BTR Model for Poisson Tensor Completion	65
F.2	Potential Applications of t-Product Geometry in Machine Learning	72

A Preliminaries and Related Works

A.1 Preliminaries of t-SVD

To model structured tensor-valued data in a way that respects both spectral locality and algebraic consistency, we build upon the algebra of t-scalars introduced in Section 2.1, and extend it to matrix- and module-level operations via the t-product. The resulting framework gives rise to a tensor analogue of singular value decomposition, known as the *tensor Singular Value Decomposition (t-SVD)* [32, 33], which forms the foundation for low-rank approximation, orthogonality, and geometric modeling in our setting.

Tensors as t-Matrices Over \mathbb{K}_c Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times c}$ be a third-order tensor. We view \mathcal{A} as a matrix over the t-scalar ring $\mathbb{K}_c := (\mathbb{R}^{1 \times 1 \times c}, +, *)$, where each entry $\mathcal{A}(i, j, :)$ is a discrete t-scalar of length c . Thus, \mathcal{A} becomes an element of $\mathbb{K}_c^{n_1 \times n_2}$, i.e., a *t-matrix*.

Given two t-matrices $\mathcal{A} \in \mathbb{K}_c^{n_1 \times n_2}$ and $\mathcal{B} \in \mathbb{K}_c^{n_2 \times n_3}$, their t-product [31, 33] is defined as:

$$\mathcal{A} * \mathcal{B} := M^{-1} (M(\mathcal{A}) \odot M(\mathcal{B})), \quad (4)$$

where M is an orthogonal transform applied along the third dimension, and the multiplication (\odot) on the right-hand side is standard matrix multiplication performed slice-wise across the c frontal slices of $M(\mathcal{A})$ and $M(\mathcal{B})$. The result $\mathcal{C} = \mathcal{A} * \mathcal{B}$ is also a t-matrix in $\mathbb{K}_c^{n_1 \times n_3}$.

This operation promotes a convolutional interpretation in the original (tube) domain, and diagonalizes computation in the transform domain. It induces an associative, unital, and distributive algebra on third-order tensors, generalizing the matrix ring structure to the spectral tensor setting.

Define the frequency-lifted tensor $\bar{\mathcal{A}} := M(\mathcal{A}) \in \mathbb{R}^{n_1 \times n_2 \times c}$ as the result of applying M along the third axis of \mathcal{A} . Then

$$\mathcal{A} * \mathcal{B} \iff \bar{\mathcal{C}}^{(k)} = \bar{\mathcal{A}}^{(k)} \cdot \bar{\mathcal{B}}^{(k)} \quad \forall k = 1, \dots, c. \quad (5)$$

This frequency-wise formulation enables efficient computation of t-products using classical linear algebra, and preserves the frequency-localized structure of tensor-valued data.

T-SVD As illustrated in Figure A.1, the t-SVD factorizes a t-matrix $\mathcal{A} \in \mathbb{K}_c^{n_1 \times n_2}$ as [31, 33]:

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \quad (6)$$

where $\mathcal{U} \in \mathbb{K}_c^{n_1 \times n_1}$ and $\mathcal{V} \in \mathbb{K}_c^{n_2 \times n_2}$ are orthogonal t-matrices, and $\mathcal{S} \in \mathbb{K}_c^{n_1 \times n_2}$ is an *f-diagonal* tensor, i.e., each frontal slice of $M(\mathcal{S})$ is diagonal [31, 33].

The t-SVD enables the definition of *tubal rank*, a measure of rank that captures frequency-structured redundancy. Specifically, the tubal rank of \mathcal{A} is the number of nonzero singular tubes of \mathcal{S} in the t-SVD formulation (6):

$$\text{rank}_t(\mathcal{A}) := \# \{i \mid \|\mathcal{S}(i, i, :)\|_2 \neq 0\}, \quad (7)$$

This generalizes the notion of matrix rank by measuring the number of active spectral components jointly across all slices. It is particularly well-suited for modeling signals with localized frequency energy or structured low-rank variation across channels [62, 61, 39, 28, 23].

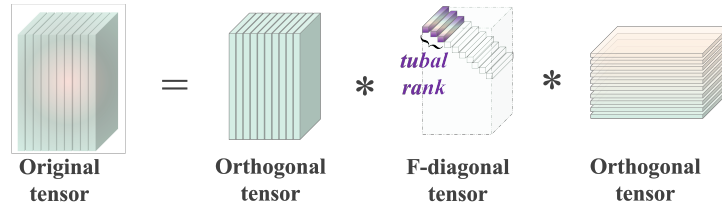


Figure A.1: Illustration of the t-SVD: the original tensor is decomposed into the product of two orthogonal tensors and an f-diagonal tensor, where the tubal rank is the number of non-zero f-diagonal tubes.

As shown in [31, 33], the t-SVD of a third-order tensor \mathcal{A} can be computed by first applying the transform M along the third mode to obtain $\bar{\mathcal{A}} := M(\mathcal{A})$. The resulting tensor consists of c frontal slices $\bar{\mathcal{A}}^{(k)} \in \mathbb{R}^{n_1 \times n_2}$ for $k = 1, \dots, c$, each of which admits a standard matrix SVD:

$$\bar{\mathcal{A}}^{(k)} = \bar{U}^{(k)} \bar{S}^{(k)} \bar{V}^{(k)\top}, \quad k = 1, \dots, c. \quad (8)$$

These components are then stacked to form $\bar{U}, \bar{S}, \bar{V}$, respectively:

$$\bar{U} = \text{stack}(\bar{U}^{(k)}), \quad \bar{S} = \text{stack}(\bar{S}^{(k)}), \quad \bar{V} = \text{stack}(\bar{V}^{(k)}),$$

and the final t-SVD factors are recovered by applying the inverse transform:

$$U = M^{-1}(\bar{U}), \quad S = M^{-1}(\bar{S}), \quad V = M^{-1}(\bar{V}).$$

The t-SVD framework is built on a *fundamental shift in modeling perspective* [32, 33, 6]: rather than viewing third-order tensors as arrays of raw entries or slices, it interprets each tube (a mode-3 fiber) as a structured algebraic entity called a *t-scalar*, which supports well-defined operations in the transform domain. This interpretation is formalized through the equivalence between the original and transformed domains in the t-product (Eqs. (4) and (5)) and in the t-SVD decomposition (Eqs. (6) and (8)):

The t-SVD philosophy treats tubes as algebraic units and operates in the frequency domain to reveal low-rank structure and coherent variation across spectral modes.

This modeling perspective leads to two essential design principles. First, by transforming tensors along the third mode via an orthogonal operator, the t-product simplifies to slice-wise matrix multiplication in the frequency domain. This diagonalization makes low-rank decomposition analytically tractable and computationally efficient. Second, by working over the t-scalar ring \mathbb{K}_c , one can naturally extend classical matrix notions such as orthogonality, rank, and projection to tensor-valued data while preserving spectral alignment and multi-channel structure. These principles have implications that go beyond numerical factorization. The t-SVD provides a canonical representation aligned with the transform structure of the data, which supports the construction of local frames, projections, and distance metrics consistent with frequency-wise organization.

In our *t-product geometry*, the t-SVD plays multiple roles: it defines canonical decompositions over t-modules, induces spectral metrics, and supports the construction of projection operators and local approximations. These properties are essential for defining orthogonality, curvature, and learning-theoretic regularity on t-manifolds. Moreover, the transform-consistent nature of t-SVD ensures that local geometric operations (e.g., gradient descent, projection) respect the frequency structure of the data, making it a natural tool for geometry-aware learning in tensor settings.

A.2 Related Work

This work integrates ideas from differential geometry, nonparametric statistics, tensor algebra, and neural network models for structured data. In what follows, we briefly review relevant contributions in each area and explain how they lay the groundwork for a geometric framework over the t-product algebra for tensor learning.

Differential Geometry and Manifold Learning Classical differential geometry provides the foundational tools for analyzing smooth structures in high-dimensional spaces [34]. Concepts such as charts, tangent spaces, derivations, and Riemannian metrics are central to understanding curvature, local linearity, and intrinsic dimensionality [36, 46]. These ideas underpin manifold learning methods that aim to uncover low-dimensional structure in data, including Laplacian Eigenmaps, Isomap, and LLE [55, 5, 46, 26].

In addition to algorithmic developments, a number of works have contributed to the theoretical formalization of the *manifold hypothesis*, which posits that real-world data distributions are often concentrated near low-dimensional smooth manifolds. For instance, [18] proposed a geometric testing framework based on interpolation and extension techniques to assess whether a given dataset is compatible with a smooth manifold structure. Building on this perspective, a series of works [74, 75, 76, 17, 16] developed manifold fitting methodologies. These approaches utilize local least

squares estimation, combined with curvature-sensitive regularity conditions, to control approximation errors and provide statistical guarantees under weak structural assumptions.

These contributions [18, 74, 75, 76, 17, 16] offer both conceptual motivation and analytical strategies for understanding whether and how high-dimensional data exhibits geometric regularity. Although these methods focus on real-valued coordinate systems in Euclidean space, their central insight that data geometry can be systematically modeled plays an important role in our formulation of tensor manifolds under the t-product algebra. In our setting, we extend this perspective to contexts where the underlying coordinate algebra is not a field but a structured ring defined via transform-based multiplication.

Nonparametric Regression on Low-Dimensional Manifolds A substantial literature addresses the problem of estimating functions defined on or near low-dimensional manifolds [29, 10, 9, 58, 81]. In particular, [29] analyze the generalization behavior of ReLU neural networks trained on data sampled from a Hölder-smooth function supported on a compact Euclidean manifold. They derive non-asymptotic risk bounds that scale with the intrinsic dimension of the manifold rather than the ambient space. Related works by [58], [10], and others study the approximation of smooth functions on manifolds using kernel regression, spectral graph methods, and local polynomial estimators.

These studies offer two key insights. First, they demonstrate that neural networks can effectively adapt to the intrinsic geometry of the input space when appropriately regularized or structured. Second, they reveal that function smoothness and manifold curvature jointly influence generalization error. While the settings considered are scalar-valued and Euclidean, the principles developed therein, such as controlling approximation error via spectral smoothness, or linking function regularity to data geometry, directly inform our analysis of function learning over tensor-valued domains governed by the t-product.

In particular, our work adopts the strategy of decomposing a function class into frequency-localized components and controlling their approximation using structured neural networks. The techniques developed in Euclidean settings serve as essential blueprints for adapting this logic to transform-domain tensor models.

Tensor Algebra and the t-Product The t-product algebra was introduced by [6] and formalized as a framework for defining multiplication between third-order tensors using circular convolution or, more generally, transform-based product [32, 31, 33]. This leads to the definition of t-scalars, t-vectors, and t-matrices, where the fundamental scalar object is a 1D tube (a vector of length c), and multiplication is performed in the transform domain (e.g., via the DFT or DCT) [31, 61, 39]. This algebra naturally gives rise to a tensor-specific version of singular value decomposition, allowing low-rank factorization via the t-SVD [32, 31, 33].

This representation has seen broad applications in tensor completion [42, 79, 62], subspace clustering [83, 68], robust PCA [23, 41], and graph signal modeling [13]. Central to many of these models is the concept of a *t-module*, which generalizes linear subspaces to tensor spaces defined over the t-product ring \mathbb{K}_c [83, 68, 30, 69]. T-modules capture global linear structure in tensor data and enable interpretable low-rank modeling.

Our work builds directly on this algebraic foundation. While previous methods emphasize algebraic efficiency and rank-based regularization, we extend the framework by equipping t-modules with differential structure: defining smoothness, tangent derivations, and metrics in a way that remains consistent with the t-product.

Geometric Modeling in Tensor Spaces There has been growing interest in introducing geometric structure into tensor representations. For instance, [19, 59] and [45] propose t-Grassmannian and t-Stiefel manifolds, generalizing classical concepts from matrix manifolds to the t-product setting. These constructions enable Riemannian optimization over tensor-valued subspaces, and have been applied to subspace estimation and dictionary learning. Other works integrate tensor models with graph geometry, leveraging t-SVD to define spectral filters and propagation operators [51, 24, 70, 44, 37].

These efforts demonstrate the promise of incorporating geometric structure into tensor algebra, with a focus on specific contexts such as Stiefel or Grassmann manifolds, or discretized graph-based models. Our work takes a complementary direction, aiming to develop a general differential

geometric framework for structured tensor spaces that supports local charts, vector fields, and smooth function spaces, while remaining compatible with the algebraic structure and spectral semantics of the t-product.

Neural Networks for Structured Tensor Learning Tensor neural networks (TNNs) incorporate t-product-based operations into deep learning architectures to preserve tensor structure during computation [48, 37, 61]. These models replace standard matrix multiplication with t-product layers, thereby maintaining spatial, spectral, or temporal organization in input data. Applications include video prediction, hyperspectral image processing, and graph-based learning [44, 70, 52, 37].

While TNNs demonstrate strong empirical performance, much of the current work focuses on architecture design and practical evaluation. Fewer studies have addressed their theoretical approximation properties, especially under geometric assumptions on the input. In this context, our theoretical analysis draws inspiration from studies of neural networks on Euclidean manifolds and aims to provide a basis for future investigation into TNNs on tensor manifolds.

Summary Taken together, these areas of research provide a well-established and conceptually complementary foundation for the modeling of high-dimensional structured data. Classical differential geometry and manifold learning offer tools for characterizing smooth variations and low-dimensional structure; nonparametric learning provides theoretical principles for function approximation on curved domains; the t-product algebra enables algebraically consistent operations on tensors with spectral structure; and tensor neural networks furnish scalable architectures tailored to such representations. *This work draws on and integrates these perspectives, taking a preliminary step toward formulating a unified geometric framework that respects both the algebraic properties and the transform-domain semantics induced by the t-product.*

B The Mathematical Foundations of t-Product Geometry

This section develops the mathematical structures underlying t-product geometry, providing the foundation for modeling structured tensor data within a differential geometric framework. The construction directly addresses **Q1** by formulating a smooth geometric theory over t-scalars that unifies discrete and continuous tensor representations. It is grounded in algebraic principles, particularly the replacement of traditional scalar fields with structured commutative rings, and draws upon tools from classical differential geometry, infinite-dimensional analysis, and algebraic geometry. This foundation enables the definition of geometric objects and operations directly on tensor data, with smoothness defined in a frequency-aware and transform-consistent manner.

B.1 From Fields to Commutative Rings: The Algebraic Scalar System

In classical differential geometry [34], the base scalars are real numbers, forming a *field* over which vector spaces and differentiable manifolds are defined. This field structure ensures invertibility (except for zero), allowing standard linear algebra tools such as bases, inner products, and coordinate changes.

In t-product geometry, the *scalar field* is replaced by a *commutative ring* \mathbb{K} equipped with a non-standard multiplication, designed to encode structured correlations such as color channels, temporal slices, or frequency components [67, 83, 68, 80, 61]. We consider two typical choices for \mathbb{K} :

- $\mathbb{K}_c = (\mathbb{R}^{1 \times 1 \times c}, +, *)$: where each element, or t-scalar, is a tube representing a finite set of structured channels (See Section 2.1 and Fig. 1-A). The multiplication $*$ is defined via a linear invertible transform M (e.g., DCT), and operates as $\mathbf{a} * \mathbf{b} = M^{-1}(M(\mathbf{a}) \odot M(\mathbf{b}))$.
- $\mathbb{K}_\infty = (C^\infty(\mathbb{R}), +, *)$: where t-scalars are smooth real-valued functions on \mathbb{R} , and multiplication is defined in a similar way using a continuous orthonormal transform (See Section 2.1 and Fig. 1-B).

The key idea is that \mathbb{K} captures not only algebraic operations but also the spectral structure of data. The t-product $*$, although associative and commutative, is defined componentwise in the transform domain, enabling the preservation of frequency-wise interactions, which is a property crucial for modern data such as images, videos, and time series [67, 83, 68, 80].

Replacing a field by a ring has several mathematical implications. First, we must use modules instead of vector spaces as the basic algebraic structure. Second, we must carefully define derivations, differentials, and metrics that respect the non-standard multiplication. These ideas parallel developments in algebraic geometry, where rings of functions (rather than fields) serve as the foundation for geometry [57, 20, 22].

B.2 From Charts to Modules: The Definition of t-Manifold

Let \mathcal{M} be a topological space modeling the domain of tensor data. According to Definition 1, a *t-manifold* of intrinsic dimension p is defined by a locally finite atlas $\{(U_\alpha, \varphi_\alpha)\}$, where each $\varphi_\alpha : U_\alpha \rightarrow V_\alpha \subseteq \mathbb{K}^p$ is a homeomorphism onto its image. Here, \mathbb{K}^p denotes the p -fold Cartesian product of the scalar ring \mathbb{K} , equipped with the product topology, and is interpreted as a free module over \mathbb{K} . This allows each chart to carry both topological and algebraic structure, reflecting the modular nature of tensor-valued representations.

To define smoothness, we impose a regularity condition on the transition maps $\varphi_\alpha \circ \varphi_\beta^{-1}$. Specifically, we require that their components become C^β -smooth after transformation by M and its inverse. In the discrete case where $\mathbb{K} = \mathbb{K}_c$, this condition reads

$$[M \circ \varphi_\alpha \circ \varphi_\beta^{-1} \circ M^{-1}]_k \in C^\beta(\mathbb{R}^p), \quad \forall k = 1, \dots, c, \quad (9)$$

and in the continuous case where $\mathbb{K} = \mathbb{K}_\infty$, we require

$$[M \circ \varphi_\alpha \circ \varphi_\beta^{-1} \circ M^{-1}](t) \in C^\beta(\mathbb{R}^p), \quad \forall t \in \mathbb{R}.$$

These smoothness conditions are imposed *componentwise and independently* across all frequency indices or parameters, meaning that each frequency slice defines its own smooth local coordinate transformation. In other words, the transform-domain representation of a t-manifold decomposes

into a family of mutually independent p -dimensional smooth manifolds—one for each frequency component—while remaining algebraically unified under the inverse transform M^{-1} .

Intuitively, a t -manifold models structured tensor data as a locally smooth object embedded in a high-dimensional tensor space, with each local chart mapping into a module over the t -scalar ring. Unlike classical manifolds that rely on real-valued coordinates, the structure here reflects both the algebraic and spectral characteristics of the t -product, where *frequency-wise independence and transform consistency together provide a principled geometric representation of tensor data*. The following remarks and examples clarify several aspects of this formulation.

Remark 2 (Uniform Chart Dimension). For simplicity, we assume that all frequency components share a common chart dimension p , so that each local chart maps into the same module \mathbb{K}^p . A more general formulation could allow each frequency index k (or parameter t in the continuous setting) to be associated with a distinct real dimension $p^{(k)}$, resulting in variable-dimensional charts across frequencies. Such a construction would require adaptive structure sheaves and frequency-dependent chart systems in Section B.3. In this work, we restrict to the uniform- p case to maintain conceptual clarity and reduce notational complexity.

Example 1 (Classical Manifold as a Special Case). When the scalar ring is chosen as $\mathbb{K} = \mathbb{R}$ and the transform is taken to be the identity operator $M = \text{Id}$, the notion of a t -manifold coincides with that of a standard C^β -smooth manifold. In this case, the t -product reduces to pointwise multiplication, and each local chart maps into the Euclidean space \mathbb{R}^p as in classical differential geometry.

Example 2 (T-Module as a Trivial t -Manifold over \mathbb{K}_c). In our framework, a t -module is defined as the free module \mathbb{K}^p over the scalar ring \mathbb{K}_c , equipped with the product topology and pointwise operations under the t -product. This space forms a *trivial t -manifold* in the sense that the entire domain is globally smooth and requires no further chart transitions. Each point in \mathbb{K}^p naturally serves as its own coordinate chart, and the transform-domain regularity is preserved throughout. This trivial case serves as the local model space for general t -manifolds, which are constructed by patching together such modules via frequency-consistent transition maps.

Remark 3 (Spectral Interpretation of Smoothness). The smoothness condition on t -manifolds is imposed in the transform domain, rather than the original data domain. This ensures that transition maps are frequency-wise smooth after applying the transform M , reflecting the regularity structure captured by the t -product algebra. This approach is closely aligned with the philosophy of the t -SVD: by treating tubes as algebraic units and diagonalizing their interactions in the frequency domain, the t -SVD reveals coherent variation across spectral modes. The t -manifold inherits this perspective by enforcing frequency-consistent smoothness in its geometric structure.

B.3 From Smooth Functions to Sheaf-Based Derivations: Differential Geometry on t -Manifolds

Building upon the definition of t -manifolds, particularly the notion of smooth structure introduced in Section 2.2, we now develop the differential geometric machinery necessary for analyzing functions and geometric structures defined on t -manifolds. Unlike classical manifolds modeled on real vector spaces, t -manifolds are constructed over free modules \mathbb{K}^p and involve a commutative ring \mathbb{K} equipped with the t -product. This algebraic shift requires a foundational rethinking of how basic geometric objects, such as derivations, differentials, and metrics, are defined and interact.

To ensure compatibility with the t -product algebra and its frequency-domain semantics, we adopt a *sheaf-theoretic and module-based formulation* [57, 20, 22]. This allows each object to retain its local geometric meaning while respecting global spectral structure. At the heart of this framework lies the *structure sheaf* $\mathcal{O}_{\mathcal{M}}$ of smooth \mathbb{K} -valued functions [57, 20]. From it, we derive the tangent module $\mathcal{T}_{\mathcal{M}}$ via \mathbb{K} -linear derivations, and construct the module of differential forms $\Omega_{\mathcal{M}}^1$ using the universal property of Kähler differentials [22].

The interplay between these three objects mirrors the classical logic of C^∞ -geometry: $\mathcal{O}_{\mathcal{M}}$ provides the space of observables, $\mathcal{T}_{\mathcal{M}}$ captures their directional variation, and $\Omega_{\mathcal{M}}^1$ serves as the dual language of measurement. In the transform domain, these structures decompose into families of familiar analytic objects, making them both geometrically principled and computationally tractable. We now proceed to define and relate these three foundational components.

B.3.1 Structure Sheaf, Tangent Derivations, and Kähler Differentials

To construct differential structures compatible with the t-product algebra, we begin by replacing the classical scalar field of smooth real-valued functions with a frequency-aware function sheaf. This allows a principled extension of vector fields and differential forms to the setting of structured tensor data. The resulting framework unifies the classical, algebraic, and frequency-domain perspectives while supporting practical modeling needs in machine learning.

Structure Sheaf $\mathcal{O}_{\mathcal{M}}$ In classical differential geometry, the basic object of smooth calculus is the sheaf $C^\infty(\mathcal{M})$ of real-valued smooth functions.

On a t-manifold, this role is played by the structure sheaf $\mathcal{O}_{\mathcal{M}}$ (Definition 3), which assigns to each open set $U \subset \mathcal{M}$ the set of \mathbb{K} -valued “smooth” functions $f : U \rightarrow \mathbb{K}$ such that their transform Mf satisfies the following frequency-wise smoothness condition:

$$\mathcal{O}_{\mathcal{M}}(U) := \left\{ f : U \rightarrow \mathbb{K} \mid \begin{cases} [Mf]_k \in C^\beta([M(U)]_k), & \text{for all } k = 1, \dots, c \quad (\text{discrete case}) \\ [Mf](t) \in C^\beta([M(U)](t)), & \text{for all } t \in \mathbb{R} \quad (\text{continuous case}) \end{cases} \right\}. \quad (10)$$

From the viewpoint of algebraic geometry, this generalizes the structure sheaf of regular functions on a scheme. The key difference lies in the scalar ring: instead of \mathbb{R} or a polynomial ring, $\mathcal{O}_{\mathcal{M}}$ takes values in a transform-defined commutative ring \mathbb{K} , often encoding spectral or temporal patterns.

In the frequency domain, the structure sheaf becomes a product of classical function spaces indexed by frequency. This makes the smoothness condition verifiable slice-wise, allowing local analysis via spectral filters or channel-specific function classes.

In the context of machine learning and signal processing, the structure sheaf $\mathcal{O}_{\mathcal{M}}$ may serve as a data-dependent function space for modeling structured signals, such as smooth video frames or band-limited hyperspectral data. While this interpretation remains theoretical at present, it provides a natural candidate domain for defining learnable functions on t-manifolds.

Remark 4 (Smooth Structure Encoded by the Sheaf). The structure sheaf $\mathcal{O}_{\mathcal{M}}$ encapsulates the smooth structure of the t-manifold in Definition 1. Rather than encoding smoothness via charts and transition maps alone, the sheaf organizes smooth \mathbb{K} -valued functions in a way that is compatible with local tensor representations and global spectral semantics. In this sense, the smooth structure of \mathcal{M} is not only topological but also algebraic, governed by the transform-defined regularity of its function spaces. This sheaf-theoretic formulation parallels the role of structure sheaves in algebraic geometry, where geometric properties are reflected in the algebra of functions.

Tangent Space (Module) $\mathcal{T}_{\mathcal{M}}$ In Definition 4, the tangent space (or tangent module) is the module of \mathbb{K} -linear derivations on $\mathcal{O}_{\mathcal{M}}$:

$$\mathcal{T}_{\mathcal{M}}(U) = \{ D : \mathcal{O}_{\mathcal{M}}(U) \rightarrow \mathcal{O}_{\mathcal{M}}(U) \mid D(f * g) = D(f) * g + f * D(g) \}.$$

This generalizes the classical notion of vector fields as directional derivatives on real-valued functions. The *Leibniz condition* remains, but scalar multiplication is now governed by the t-product [57], making derivations sensitive to spectral structure.

In algebraic geometry, derivations define the tangent sheaf $\mathcal{T}_{\mathcal{M}}$ of a scheme or ringed space. Our version adapts this to non-field base rings, necessitating a sheaf of \mathbb{K} -modules rather than vector bundles.

In the transform domain, each derivation decomposes into a family of classical vector fields $\{X^{(k)}\}$ or $\{X^{(t)}\}$ acting slice-wise on $[Mf]_k$ or $[Mf](t)$. Thus,

$$D(f) = M^{-1} \left(\{X^{(k)}([Mf]_k)\} \right).$$

This structure preserves classical intuition at each frequency, while enforcing global algebraic consistency.

In machine learning, elements of $\mathcal{T}_{\mathcal{M}}$ can be viewed as describing local variations of tensor-valued functions. For example, they may reflect directions along which pixel intensities change across color channels or spectral bands. While such interpretations are preliminary, they point to a possible connection between the tangent module and structured variation in data.

Differential 1-Forms $\Omega_{\mathcal{M}}^1$ In Definition 5, differential 1-forms are the *Kähler differentials* [20, 57] over $\mathcal{O}_{\mathcal{M}}$. Each differential df is a formal symbol satisfying:

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

and

$$df(X) = X(f), \quad \forall X \in \mathcal{T}_{\mathcal{M}}(U).$$

The module $\Omega_{\mathcal{M}}^1$ is constructed so that it is dual to $\mathcal{T}_{\mathcal{M}}$ under evaluation [57]—just as in classical differential geometry, where 1-forms are linear functionals on tangent vectors [34].

In algebraic geometry [22, 20], this is the standard approach to defining differentials in a coordinate-free, module-theoretic manner. It remains valid over general commutative rings and supports the exterior algebra needed for higher-order forms.

In the transform domain, each differential 1-form df corresponds to a collection of slice-wise classical differentials $d([Mf]_k)$, which enables the extension of exterior calculus to individual spectral components. The pairing $df(X)$ is computed frequency-by-frequency and subsequently mapped back via inverse transform to remain within \mathbb{K} , in a manner consistent with the structural principles underlying the t-SVD.

In machine learning, 1-forms serve as interpretable descriptors of local change, enabling spectral sensitivity analysis and Jacobian-based interpretability. They also underpin manifold-based reconstruction techniques where smoothness or sparsity of derivatives in certain directions is enforced.

The Relationship Among $\mathcal{O}_{\mathcal{M}}$, $\mathcal{T}_{\mathcal{M}}$ and $\Omega_{\mathcal{M}}^1$ The structure sheaf $\mathcal{O}_{\mathcal{M}}$, the tangent module $\mathcal{T}_{\mathcal{M}}$, and the differential module $\Omega_{\mathcal{M}}^1$ together form a fundamental triple in the differential geometry of t-manifolds. The sheaf provides the function space, the tangent module captures directions of infinitesimal variation, and the module of 1-forms measures directional change along these directions. These objects satisfy a natural duality relation:

$$\Omega_{\mathcal{M}}^1(U) = \text{Hom}_{\mathcal{O}_{\mathcal{M}}(U)}(\mathcal{T}_{\mathcal{M}}(U), \mathcal{O}_{\mathcal{M}}(U)),$$

where $\text{Hom}_{\mathcal{O}_{\mathcal{M}}(U)}$ denotes the set of $\mathcal{O}_{\mathcal{M}}(U)$ -linear maps from the tangent module to the structure sheaf. That is, each differential 1-form on U acts as a “linear functional” on vector fields, with compatibility defined over the t-scalar-valued function ring.

Remark 5 (Alternative Construction via Jet Modules). While we define the tangent module $\mathcal{T}_{\mathcal{M}}$ and differential module $\Omega_{\mathcal{M}}^1$ using derivations and Kähler differentials over the structure sheaf $\mathcal{O}_{\mathcal{M}}$, these objects may also be constructed via first-order jet modules [57]. In this alternative formulation, vector fields correspond to equivalence classes of first-order function variations, and 1-forms arise as universal derivations. We do not pursue this approach here, as the sheaf-theoretic formalism aligns more naturally with the frequency-domain structure induced by the t-product.

This algebraic formulation forms the starting point for constructing Riemannian metrics, affine connections, and geometric operators in the t-product setting.

B.3.2 Riemannian Geometry on t-Manifolds

Riemannian Metric By Definition 6, a Riemannian metric on a t-manifold is a symmetric, positive-definite \mathbb{K} -bilinear map

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

which assigns, for each pair of vector fields, a \mathbb{K} -valued smooth function. In classical differential geometry, the metric measures inner products between real-valued tangent vectors and induces a local notion of length, angle, and volume [56]. The generalization to \mathbb{K} -modules requires bilinearity with respect to the t-product and evaluation within the transform-consistent structure sheaf.

From the perspective of algebraic geometry, such a metric corresponds to a symmetric $\mathcal{O}_{\mathcal{M}}$ -bilinear form on a coherent sheaf, analogous to defining a global section of $\text{Sym}^2 \mathcal{T}_{\mathcal{M}}^*$. The distinction here lies in the underlying scalars: instead of polynomials or regular functions, the base ring is frequency-aware and built from structured signal components.

In the transform domain, the metric decomposes into a set of slice-wise inner products:

$$[Mg(X, X)]_k > 0 \quad (\text{discrete case}), \quad [Mg(X, X)](t) > 0 \quad (\text{continuous case}).$$

This means that, for each frequency component, the inner product reduces to a classical one defined on the corresponding slice. Such a decomposition allows the geometric structure to remain compatible with frequency-localized variation, as commonly observed in settings such as color channels in images or band-limited features in audio signals. While these interpretations are illustrative, they should be regarded as preliminary and primarily serve to motivate the frequency-wise formulation.

In the context of machine learning and signal processing, such a metric provides a natural framework for *expressing similarity between structured data points*, such as image patches or video segments. It offers a frequency-aware notion of distance that can reflect spectral heterogeneity in the data, which may be more appropriate than standard Euclidean metrics in certain settings. While these ideas suggest potential relevance to tasks such as clustering, manifold regularization, and feature extraction [5, 46], their practical implications remain to be fully explored.

Gradient and Laplacian The gradient of a function $f \in \mathcal{O}_{\mathcal{M}}(U)$ in Definition 7 is the t-vector field $\nabla f \in \mathcal{T}_{\mathcal{M}}(U)$ satisfying

$$g(\nabla f, X) = df(X), \quad \forall X \in \mathcal{T}_{\mathcal{M}}(U).$$

This is the natural generalization of the classical gradient via the Riesz representation theorem, reinterpreted in a module-theoretic context. In algebraic geometry, this operation corresponds to identifying derivations with differentials through a metric-induced isomorphism between $\mathcal{T}_{\mathcal{M}}$ and its dual.

The Laplace operator Δf is then defined in Definition 8 as the divergence of the gradient:

$$\Delta f = \text{div}(\nabla f).$$

The divergence operator is constructed using the volume form induced by the metric and is interpreted transform-wise via classical divergence in each spectral component.

This construction gives rise to frequency-aware diffusion operators, which extend classical notions of the Laplacian to the setting of t-manifolds. In the context of machine learning, such operators may be used to define spectral filters, denoising objectives, or regularization terms that promote smoothness in dominant spectral components. This perspective offers a geometric foundation for frequency-informed modeling and generalizes classical Laplacian regularization techniques used in semi-supervised learning and spectral embedding. In particular, it provides a principled extension of spectral methods used in t-SVD-based graph models [13, 51, 44, 70, 37], where our formulation endows such models with intrinsic, manifold-consistent Laplacians. These constructions remain theoretical, but they suggest a pathway toward incorporating geometric structure into tensor-based learning pipelines.

Affine Connection A connection [57] on the tangent module is a \mathbb{K} -bilinear operator

$$\nabla : \mathcal{T}_{\mathcal{M}}(U) \times \mathcal{T}_{\mathcal{M}}(U) \rightarrow \mathcal{T}_{\mathcal{M}}(U),$$

which satisfies the $\mathcal{O}_{\mathcal{M}}$ -linearity condition

$$\nabla_X(f * Y) = X(f) * Y + f * \nabla_X Y,$$

for all $f \in \mathcal{O}_{\mathcal{M}}(U)$. This generalizes directional differentiation of vector fields and enables parallel transport and curvature analysis on the t-manifold.

In classical differential geometry, such a connection defines how vectors “move” along curves. Algebraically, it is a rule for differentiating sections of a sheaf of modules. In the frequency domain, it induces a family of classical connections indexed by frequency or transform parameter:

$$[M(\nabla_X Y)]_k = \nabla_{X^{(k)}}^{(k)} Y^{(k)} \quad \text{or} \quad [M(\nabla_X Y)](t) = \nabla_{X^{(t)}}^{(t)} Y^{(t)}.$$

This formulation preserves the independence of spectral components and ensures that geodesic flows respect the internal structure of the data.

Christoffel Symbols and Levi-Civita Connection In a local \mathbb{K} -module basis $\left\{ \frac{\partial}{\partial p_i} \right\}$, the connection is expressed via Christoffel symbols:

$$\nabla_{\frac{\partial}{\partial p_i}} \left(\frac{\partial}{\partial p_j} \right) = \sum_k \Gamma_{ij}^k * \frac{\partial}{\partial p_k}, \quad \Gamma_{ij}^k \in \mathcal{O}_{\mathcal{M}}(U).$$

The Levi-Civita connection on t-Manifold in Definition 9 is the unique torsion-free, metric-compatible connection determined by the metric:

$$\nabla_X Y - \nabla_Y X = [X, Y], \quad X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

These conditions are imposed slice-wise in the transform domain, resulting in Christoffel symbols that take the form of structured, frequency-dependent functions. These symbols encode how directional derivatives interact within each spectral mode, preserving alignment with the underlying tensor structure.

In the context of machine learning, such a connection may inform the design of smooth interpolation procedures (e.g., along geodesics), curvature-aware regularization terms, or data-driven transport models. It may also motivate the exploration of Riemannian optimization techniques, where update directions are adapted to the geometry of the underlying t-manifold. While these interpretations are preliminary, they highlight possible ways in which geometric structure could be incorporated into learning frameworks for structured data.

Geodesics Given the Levi-Civita connection, geodesics are further defined in Definition 9 as curves $\gamma(t)$ satisfying:

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0.$$

Locally, this gives the geodesic equation:

$$\frac{d^2 p^k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k * \frac{dp^i}{dt} \frac{dp^j}{dt} = 0.$$

Each coordinate component satisfies a second-order differential equation whose coefficients are frequency-structured. The solution to this equation traces the shortest path with respect to the t-metric and serves as the basis for exponential maps, normal coordinates, and interpolation algorithms.

Geodesics can be used to model gradual transitions in structured data, such as appearance changes or motion patterns in video sequences. In learning contexts, they may provide a tool for latent space interpolation or regularization in autoencoder architectures. These applications remain exploratory, but highlight the potential relevance of geodesic structure in modeling smooth variations within high-dimensional data.

B.3.3 Summary and Potential Extensions

The geometric framework developed above provides a transform-consistent extension of classical differential geometry to the setting of tensor-valued data. By systematically reformulating fundamental geometric objects—including functions, vector fields, differentials, metrics, connections, and geodesics—within the language of \mathbb{K} -modules and their behavior under orthogonal transforms, the resulting theory remains algebraically coherent while reflecting the structured nature of real-world data.

The proposed framework admits several natural extensions, including the following:

- **Curvature tensors:** One can define curvature quantities such as the Riemann tensor $R(X, Y)Z$, Ricci curvature, and scalar curvature by computing commutators of covariant derivatives [57].
- **Geodesic-based structures:** Geodesic variation can be studied via Jacobi fields, which characterize how geodesics respond to perturbations. In parallel, exponential maps provide a means for constructing local normal coordinates by tracing geodesics from a given point [1].
- **Stochastic geometry:** Brownian motion and related stochastic processes could be generalized to the t-manifold setting via the Laplace–Beltrami operator, offering a pathway toward probabilistic modeling and spectral learning [46].

Although these extensions are mathematically natural and structurally compatible with the present formulation, we do not pursue them in this work. Our objective is to provide a minimal yet self-contained differential-geometric foundation that supports the representation, approximation, and optimization tasks developed in the main text.

C Theoretical Analysis For t-Manifold Hypothesis Testing

This appendix provides a statistically grounded answer to *the first aspect* of **Q2**:

Can we decide, from a finite number of samples, whether tensor-valued data in \mathbb{K}_c^d lies close to a low-dimensional smooth t-manifold?

We extend the manifold hypothesis testing framework of Fefferman et al. [18] to the t-product space \mathbb{K}_c^d , developing a statistical procedure that certifies proximity to a structured t-manifold class under mild geometric assumptions. Throughout, tensor samples $\{x_i\}_{i=1}^n$ are assumed to lie in the unit ball $\mathbb{B}_{\mathbb{K}_c^d}(0, 1) \subset \mathbb{K}_c^d$.

To support the theoretical framework in Section 3.1, we formalize the local and global geometry of admissible t-manifolds in \mathbb{K}_c^d , including local $C^{1,1}$ graph representations, compatibility across overlapping patches, and the extension of discretized anchor patch sets into smooth global structures. We then quantify the complexity of the associated squared-distance function class and establish uniform convergence guarantees over $\mathcal{G}(p, V, \tau)$ with an explicit sample complexity bound.

The construction of such manifolds from data (manifold fitting) will be presented in Section D, and the learning of functions near these manifolds in Section E. We use C, C_1, C', c', c_1 , etc., to denote positive constants that may vary from line to line and may depend on the channel number c and the intrinsic dimension p .

C.1 General Strategy of the Proof

The proof of Theorem 1 closely follows the geometric–analytic framework developed by Fefferman et al. [18], with necessary adaptations to the transform-domain module structure of \mathbb{K}_c^d . Our goal is limited to verifying that the essential geometric and statistical arguments used in the Euclidean setting continue to hold when extended slice-wise under the t-product representation. Specifically, we aim to determine, from n i.i.d. t-vector samples, whether the underlying distribution is concentrated in a small neighborhood of some p -dimensional t-manifold $\mathcal{M} \in \mathcal{G}(p, V, \tau)$, or instead remains uniformly separated from all such manifolds within that class.

1. *Geometric formulation.* The admissible class $\mathcal{G}(p, V, \tau)$ is described through its frequency-domain slices $\mathcal{M}_k \subset \mathbb{R}^d$, each a p -dimensional $C^{1,1}$ manifold with bounded volume and reach. The global manifold \mathcal{M} is obtained by inverse orthogonal transform, which preserves the relevant metric and smoothness properties across slices.
2. *Manifold construction.* We follow the local-to-global construction scheme of [18]: local $C^{1,1}$ graphs are built around geometrically regular anchor patches on each slice and then glued together in a transform-consistent manner. This produces a globally smooth candidate manifold with controlled curvature and reach in the tensor domain.
3. *Sample complexity and uniform approximation.* The empirical-process and uniform-convergence results established in [18] for Euclidean manifolds are applied independently to each frequency slice. Because the orthogonal transform M preserves distances and differentiability, the order of the sample complexity extends to the tensor setting up to a constant factor depending on the number of frequency components c . We do not attempt to refine or re-derive the constants involved, but rely on the structure of the original arguments.
4. *Empirical risk and decision rule.* A finite hypothesis family is obtained by discretizing the geometric parameters, including anchor locations and tangent directions, at a discrete scale determined by the lower bound τ on the spectral reach. Empirical risk minimization over this finite family yields the test statistic: a small empirical squared distance indicates that the data lie close to some member of $\mathcal{G}(p, V, \tau)$, whereas a large distance excludes all admissible candidates.

C.2 Geometric and Statistical Lemmas for t-Manifolds

This subsection collects the auxiliary lemmas that support the proof of Theorem 1. The arguments extend the geometric and statistical components of the Euclidean framework [18] to the transform-domain representation of t-manifolds in \mathbb{K}_c^d . The lemmas verify that key properties including local $C^{1,1}$ regularity, stability under perturbations, the existence of approximate squared distance functions,

and uniform empirical convergence, continue to hold under the slice-wise structure induced by the t-product. Together they provide the geometric consistency and probabilistic control required for the main theorem.

Infinitesimal and Spectral Reach To ensure that the underlying manifolds are well-behaved geometrically, we follow the setup of Fefferman et al. [18] and specify the regularity assumptions imposed on t-manifolds and their frequency-domain slices. Ref. [18] assumes that each manifold has *infinitesimal spectral reach* at least one and *global spectral reach* $\tau \leq 1$. The infinitesimal reach lower bound ensures locally unique normal projections and bounded curvature, while the normalization $\tau \leq 1$ fixes the geometric scale for analysis.

In our tensor setting, each frequency slice $\mathcal{M}_k \subset \mathbb{R}^d$ inherits these conditions:

$$\text{reach}_{\mathbb{R}^d}(\mathcal{M}_k) \geq \tau, \quad \text{infinitesimal reach}_{\mathbb{R}^d}(\mathcal{M}_k) \geq 1, \quad \tau \leq 1.$$

The first inequality guarantees slice-wise $C^{1,1}$ regularity and curvature control of order $O(1/\tau)$, while the second acts as a normalization. Under orthogonal transforms M defining the t-product, these properties are preserved in the tensor domain.

Therefore, all t-manifolds $\mathcal{M} \in \mathbb{B}_{\mathbb{K}_c^d}(0, 1) \subset \mathbb{K}_c^d$ considered in this work are assumed to satisfy the same regularity conditions of their frequency slices as in Fefferman et al. [18], ensuring well-defined tangent spaces, stable local charts, and uniform curvature bounds across frequency components.

Local Graph Representation of t-Manifolds To describe the local geometry of a t-manifold in a tractable form, it is useful to express each small neighborhood as a smooth graph over its tangent subspace. This representation makes curvature and regularity explicit and allows geometric properties to be analyzed slice by slice in the frequency domain. The following lemma extends the classical local graph representation of smooth manifolds (see [18, Claim 1]) to the tensor setting. It shows that every point on a t-manifold admits a neighborhood which, after an orthogonal transform to the frequency domain, can be written as a $C^{1,1}$ graph whose curvature is controlled by the spectral reach.

Lemma 1 (Local Graph Representation of t-Manifolds). *Let $\mathcal{M} \in \mathcal{G}(p, V, \tau)$. Then, for every point $x \in \mathcal{M}$, there exists a neighborhood $U_x \subset \mathbb{K}_c^d$ and a \mathbb{K}_c -valued $C^{1,1}$ mapping $F_{x,U_x} : \Pi_x(U_x) \rightarrow \Pi_x^\perp$ such that*

$$\mathcal{M} \cap U_x = \{y + F_{x,U_x}(y) : y \in \Pi_x(U_x)\}, \quad (11)$$

where Π_x is a p -dimensional submodule serving as the tangent approximation at x . For each frequency slice k , the component $[F_{x,U_x}]_k$ satisfies $\text{Lip}(\nabla[F_{x,U_x}]_k) \leq C/\tau$.

Proof Sketch. The proof follows the argument of [18, Claim 1], extended to the slice-wise tensor setting. Let M denote the orthogonal transform defining the t-product, and write $\hat{x} = Mx \in \mathbb{R}^{c \times d}$. For $\mathcal{M} \in \mathcal{G}(p, V, \tau)$, each slice $\mathcal{M}_k := \{[My]_k : y \in \mathcal{M}\} \subset \mathbb{R}^d$ is a p -dimensional C^2 manifold with reach at least τ . By [18, Claim 1], for each $\hat{x}_k = [Mx]_k$ there exist local coordinates $f_k : \mathbb{R}^p \rightarrow \mathbb{R}^{d-p}$ such that

$$\mathcal{M}_k \cap U_k = \{y_k + f_k(y_k) : y_k \in \pi_k(U_k)\}, \quad \text{Lip}(\nabla f_k) \leq C/\tau.$$

Define the slice-wise aggregate

$$\hat{F}(y_1, \dots, y_c) = (y_1 + f_1(y_1), \dots, y_c + f_c(y_c)),$$

and lift it back to \mathbb{K}_c^d via the inverse transform:

$$F_{x,U_x}(y) = M^{-1}(\hat{F}(My) - My), \quad U_x := M^{-1} \left(\bigtimes_{k=1}^c U_k \right).$$

Because M is orthogonal, Lipschitz and curvature bounds are preserved under this transformation. Hence F_{x,U_x} is $C^{1,1}$ with $\text{Lip}(\nabla F_{x,U_x}) \leq C/\tau$ in every spectral component.

This yields the desired local representation $\mathcal{M} \cap U_x = \{y + F_{x,U_x}(y) : y \in \Pi_x(U_x)\}$ in \mathbb{K}_c^d . \square

Discrete Approximation via Anchor Patch Sets To handle t-manifolds computationally, a continuous geometric object must be replaced by a finite and well-controlled representation. The idea is to approximate the smooth structure of the manifold by a finite collection of local charts that capture its geometry up to the desired resolution. Building upon Lemma 1, which ensures that every point on a t-manifold admits a uniformly regular $C^{1,1}$ local graph in the frequency domain, we construct such a discrete representation through a set of *anchor patches*. Each anchor patch describes a local geometric neighborhood, and their centers serve as the basic reference points for later manifold reconstruction and statistical testing.

Following the “cylinder packet” construction [18], we discretize each frequency slice of the t-manifold by selecting a finite family of $C^{1,1}$ patches with bounded radii and controlled $\dot{C}^{1,1}$ seminorms. Each patch is expressed as a $C^{1,1}$ graph over a p -dimensional tangent plane in the frequency domain. The collection of patch centers, called the *anchor patch set*, forms a well-separated, jointly covering scaffold of the t-manifold in spectral space.

Lemma 2 (Existence of anchor patch sets on t-manifolds). *Let $\mathcal{M} \in \mathcal{G}(p, V, \tau)$ be a p -dimensional t-manifold with spectral volume at most V and spectral reach at least τ . Then there exist a finite subset $X = \{x_1, \dots, x_N\} \subset \mathcal{M}$ (the anchor patch set) and a collection of local descriptors $\vec{\eta} = \{(r_i, A_i, \Pi_i)\}_{i=1}^N$ satisfying the following properties:*

1. *Slice-wise separation. For every frequency component $k \in [c]$, the projections $\{[Mx_i]_k\}$ are pairwise separated by at least τ/C in \mathbb{R}^d .*
2. *Slice-wise coverage. For each slice k , the projected centers $\{[Mx_i]_k\}$ form a τ/C -net covering the slice manifold $\mathcal{M}_k := \{[Mx]_k : x \in \mathcal{M}\} \subset \mathbb{R}^d$.*
3. *Local patch regularity. For each $x_i \in X$ and each slice k , there exists a local $C^{1,1}$ graph $\{y_k + f_{i,k}(y_k) : y_k \in \Pi_{i,k}\}$ over a p -plane $\Pi_{i,k}$, with radius $r_i \geq \tau/C$ and $\dot{C}^{1,1}$ seminorm $\|\nabla f_{i,k}\|_{\text{Lip}} \leq A_i \leq C/\tau$.*

The descriptor $\vec{\eta}$ thus records, for each anchor, its patch radius r_i , curvature bound A_i (quantifying second-order variation), and the slice-wise tangent planes $\Pi_i = \{\Pi_{i,k}\}_{k=1}^c$ that jointly characterize its local spectral geometry.

Proof Sketch. We adapt the geometric discretization argument of Fefferman et al. [18] to the t-product framework. For each frequency slice $\mathcal{M}_k \subset \mathbb{R}^d$, the reach bound τ guarantees the existence of a τ/C -net $X_k = \{z_{k,1}, \dots, z_{k,N_k}\} \subset \mathcal{M}_k$ with pairwise separation at least $\tau/(2C)$ and cardinality $N_k \leq CV_k/\tau^p$, where $\sum_k V_k \leq V$.

Define

$$\hat{X} = \bigtimes_{k=1}^c X_k, \quad X = \{M^{-1}(z_{1,j_1}, \dots, z_{c,j_c}) : z_{k,j_k} \in X_k\}.$$

Because M is orthogonal, the set X inherits both separation and coverage properties, and its size satisfies $|X| \leq CV/\tau^p$. Finally, by Lemma 1, each $x_i \in X$ admits a slice-wise $C^{1,1}$ chart with radius and $\dot{C}^{1,1}$ bounds within the prescribed limits, completing the construction. \square

Having established the existence of well-behaved anchor patch sets on smooth t-manifolds, we next examine their *stability*, that is, how such geometric scaffolds respond to small perturbations of the underlying manifold or its sample points. This step is crucial for the later stages of statistical testing and manifold fitting, where empirical observations or estimated anchor centers inevitably deviate from their true geometric locations. Intuitively, if each anchor is slightly displaced within the frequency domain while preserving relative distances and tangent orientations, the resulting perturbed configuration should still define a valid anchor patch set, possibly with slightly altered parameters. The following lemma formalizes this stability property.

Lemma 3 (Stability of anchor patch sets under slice-wise perturbations). *Let $X = \{x_i\}_{i=1}^N \subset \mathbb{K}_c^d$ be an anchor patch set associated with a t-manifold $\mathcal{M} \in \mathcal{G}(p, V, \tau)$ and local descriptors $\vec{\eta} = \{(r_i, A_i, \Pi_i)\}_{i=1}^N$, where $\Pi_i = \{\Pi_{i,k}\}_{k=1}^c$ denote the slice-wise tangent planes. Assume that X satisfies the slice-wise geometric conditions of Lemma 2 (separation, coverage, and local $C^{1,1}$ regularity) in the transform domain $M : \mathbb{K}_c^d \rightarrow \mathbb{R}^{c \times d}$.*

Consider a perturbed configuration

$$\tilde{X} = \{\tilde{x}_i\}_{i=1}^N, \quad \tilde{x}_i := (1 - c_1)(x_i + v_i), \quad \| [Mv_i]_k \|_2 < c_2 \text{ for all } k \in [c],$$

where $c_1, c_2 > 0$ are small perturbation parameters. Then, for sufficiently small (c_1, c_2) , the perturbed set \tilde{X} remains an anchor patch set of the same cardinality, with updated parameters $\tau' \geq \tau(1 - c_3)$, and descriptor

$$\vec{\eta}' = \{(r'_i, A'_i, \Pi'_i)\}_{i=1}^N, \quad r'_i = r_i(1 + O(c_1 + c_2)), \quad A'_i = (I + \Delta A_i)A_i, \quad \Pi'_{i,k} = \Pi_{i,k} + \Delta \Pi_{i,k},$$

where $\|\Delta A_i\|, \|\Delta \Pi_{i,k}\| = O(c_1 + c_2)$ and $c_3 \rightarrow 0$ as $c_1, c_2 \rightarrow 0$.

Proof Sketch. The argument relies on the local regularity and reach control of t-manifolds and follows the perturbation framework developed by Fefferman et al. [18, Lemmas 25-28]. Because the transform M is orthogonal, a perturbation $\|[Mv_i]_k\|_2 < c_1$ preserves Euclidean distances and angles within $O(c_1)$ at each frequency slice. Hence the separation and coverage properties of X carry over to \tilde{X} with parameters scaled by $(1 + O(c_1 + c_2))$.

Locally, each patch in Lemma 2 admits a $C^{1,1}$ graph representation with $\dot{C}^{1,1}$ seminorm $\leq C/\tau$. Lemmas 26 in [18] shows that small displacements of the centers or tangent planes of such graphs preserve their $C^{1,1}$ structure and reach, with curvature bounds perturbed only by $O(c_1 + c_2)$. Applying this slice-wise yields new graphs with radii $r'_i = r_i(1 + O(c_1 + c_2))$ and tangent perturbations $\|\Delta \Pi_{i,k}\| = O(c_1 + c_2)$. The reach behaves continuously under such deformations [18, Lemmas 28], so that $\tau' \geq \tau(1 - c_3)$ for a small constant c_3 .

Therefore, \tilde{X} satisfies all defining properties of an anchor patch set, with updated parameters depending smoothly on the perturbation magnitude. \square

From Anchor Patches to t-Cylinder Packets While the anchor patch set in Lemma 2 captures the intrinsic local geometry of a true t-manifold, the subsequent algorithmic construction requires a separate, parameterized family of geometric primitives that can be freely adjusted and combined to approximate unknown manifolds. In the spirit of Fefferman et al. [18], we introduce *t-cylinder packets* as finite collections of slice-wise product cylinders in the transform domain, which serve as building blocks for defining approximate squared distance functions (ASDFs).

Formally, each t-cylinder packet $\hat{O} = \{\widehat{\text{Cyl}}_i\}_{i=1}^{\tilde{N}}$ consists of product cylinders $\widehat{\text{Cyl}}_i = \times_{k=1}^c \text{Cyl}_{i,k}$, with $\text{Cyl}_{i,k} \cong 2\bar{\tau}(B_p \times B_{d-p}) \subset \mathbb{R}^d$ where $\bar{\tau} = c_4\tau$. Intuitively, a t-cylinder can be viewed as a small, tube-like neighborhood around a flat p -dimensional slice in the frequency domain. Its central section plays the role of a local tangent patch, while its surrounding volume provides the region within which nearby data points may lie. These cylinders serve as adjustable geometric templates: by varying their positions, orientations, and curvature parameters, one can span a rich family of possible local manifold shapes. Unlike anchor patches, which are extracted from the true manifold to describe its observed geometry, t-cylinder packets are synthetic constructions that define the search space of all admissible local geometries. They thus act as geometric “atoms” from which candidate manifolds are assembled. On every $\text{Cyl}_{i,k}$ we define a local quadratic model approximating the squared distance to its central section, and glue these models via a smooth partition of unity to produce the slice-wise ASDF F_k defined below.

Approximate Squared Distance Functions and t-Manifold Reconstruction To transform the discrete geometric scaffold provided by the t-cylinder packet into a smooth manifold, we require a continuous function that encodes distances to the local tangent structures. This leads to the construction of an *approximate squared distance function* (ASDF) [18, Definition 16], a smooth potential whose shape captures the geometry of the manifold to be recovered.

Given a fixed t-cylinder packet, we define a scalar field for each frequency slice whose level sets interpolate the tangent patches represented by the cylinders. In geometric terms, this field behaves like an energy landscape: regions near the manifold correspond to valleys, while the manifold itself appears as the ridge where the gradient vanishes. Each slice-wise ASDF F_k thus acts as a smooth distance potential: *its level sets follow the local geometry of \mathcal{M}_k , its gradient aligns with the normal directions, and its zero-gradient set reconstructs the manifold.* In the t-product setting, these

slice-wise potentials are then coherently combined through the inverse transform to form a global \mathbb{K}_c -smooth function on \mathbb{K}_c^d .

The following lemma formalizes the construction of the *t-approximate squared distance function* (t-ASDF) and the associated geometric objects that arise from it. Intuitively, these constructions transform the local information encoded by the t-cylinder packet into a smooth manifold approximation equipped with a well-defined normal structure. The main objects are summarized as follows:

1. *Slice-wise potentials* F_k . For each frequency slice \mathcal{M}_k , we build a smooth function F_k whose values approximate the squared distance to \mathcal{M}_k within a small tubular neighborhood $\mathcal{M}_{k,\bar{\tau}}$. The gradient and Hessian of F_k capture the local normal and curvature information, respectively. Rescaling F_k shows that its derivatives are uniformly bounded, reflecting controlled geometric regularity.
2. *Putative slice manifolds* $\widehat{\mathcal{M}}_{\text{put}}^{(k)}$. The zero set of the projected gradient $\Pi_{hi}^{(k)}(y) \nabla F_k(y) = 0$ defines a smooth submanifold that approximates \mathcal{M}_k . Geometrically, this is the locus where the potential F_k is “flat” along the tangent directions and steepest along the normals. The resulting $\widehat{\mathcal{M}}_{\text{put}}^{(k)}$ is a $C^{1,1}$ manifold close to \mathcal{M}_k both in reach and in Hausdorff distance.
3. *Global reconstruction* \mathcal{M}_{put} . Taking the product of all slice-wise manifolds and mapping back through the inverse transform M^{-1} yields a global t-manifold $\mathcal{M}_{\text{put}} \subset \mathbb{K}_c^d$. This object represents the reconstructed version of the original manifold \mathcal{M} , consistent across spectral components and smooth in the t-product sense.
4. *Normal disc bundle*. Around each point $\hat{z} \in \widehat{\mathcal{M}}_{\text{put}}$, the top $(d-p)$ eigenspace of $\nabla^2 F_k(\hat{z}^{(k)})$ defines the local normal directions in the frequency domain slice-wisely. Taking small discs in these directions gives the slice-wise normal discs $\widehat{D}_{\hat{z},k}^{\text{norm}}$, which together form a $C^{1,1}$ disc bundle in the transform domain. Applying the inverse transform yields the lifted t-normal disc bundle D_z^{norm} over \mathcal{M}_{put} , describing a smooth tubular neighborhood of the reconstructed manifold.

Together, these constructions establish a smooth correspondence between the discrete cylinder-based representation and a continuous t-manifold equipped with a well-defined normal geometry and curvature control.

Lemma 4 (t-Approximate Squared Distance Function). *There exist controlled constants $c_4 \in (0, 1)$ and $C_*, c_* > 0$ depending on c, p and absolute numerical parameters, and slice-wise functions*

$$F_k : \mathcal{M}_{k,\bar{\tau}} \rightarrow \mathbb{R}, \quad \bar{\tau} := c_4 \tau, \quad \mathcal{M}_{k,\bar{\tau}} := \{y \in \mathbb{R}^d : \text{dist}(y, \mathcal{M}_k) < \bar{\tau}\},$$

constructed by the above partition-of-unity scheme, such that the following hold:

1. (ASDF properties, slice-wise) *For each k , define the rescaled function $\widehat{F}_k(w) := F_k(z + \bar{\tau}w)/\bar{\tau}^2$ on $B_d(0, 1)$ around any $z \in \mathcal{M}_k$. Then \widehat{F}_k has uniformly bounded derivatives up to the required order, and satisfies the standard ASDF structural conditions used by Fefferman: there exist controlled c_5, C_0, C_1 so that*

$$\|\partial^\alpha \widehat{F}_k\|_{L^\infty(B_d)} \leq C_0 (|\alpha| \leq 2), \quad c_5(|y_\perp|^2 + \rho^2) \leq \widehat{F}_k(y) + \rho^2 \leq C_1(|y_\perp|^2 + \rho^2),$$

for some controlled $\rho \in (0, 1)$, where y_\perp denotes the normal component to the tangent p -plane at z . Hence $\nabla^2 F_k$ is uniformly bounded on $\mathcal{M}_{k,\bar{\tau}}$.

2. (Putative slice manifolds) *Let $\Pi_{hi}^{(k)}(y)$ denote the orthogonal projector onto the span of the top $(d-p)$ eigenvectors of $\nabla^2 F_k(y)$. Define the putative slice*

$$\widehat{\mathcal{M}}_{\text{put}}^{(k)} := \{y^{(k)} \in \mathcal{M}_{k,\bar{\tau}} : \Pi_{hi}^{(k)}(y^{(k)}) \nabla F_k(y^{(k)}) = 0\}.$$

Then $\widehat{\mathcal{M}}_{\text{put}}^{(k)}$ is a p -dimensional $C^{1,1}$ submanifold of \mathbb{R}^d , with

$$\text{reach}(\widehat{\mathcal{M}}_{\text{put}}^{(k)}) \geq c_* \tau, \quad d_H(\widehat{\mathcal{M}}_{\text{put}}^{(k)}, \mathcal{M}_k) \leq C_* \bar{\tau}.$$

Locally $\widehat{\mathcal{M}}_{\text{put}}^{(k)}$ is a $C^{1,1}$ graph over the tangent p -plane with seminorm $\leq C_/\tau$.*

3. (Product and lift) Set $\widehat{\mathcal{M}}_{\text{put}} := \times_{k=1}^c \widehat{\mathcal{M}}_{\text{put}}^{(k)} \subset \mathbb{R}^{c \times d}$ and define

$$\mathcal{M}_{\text{put}} := M^{-1}(\widehat{\mathcal{M}}_{\text{put}}) \subset \mathbb{K}_c^d.$$

Then \mathcal{M}_{put} is a p -dimensional $C^{1,1}$ t -manifold with

$$\text{reach}_{\text{spec}}(\mathcal{M}_{\text{put}}) \geq c_* \tau, \quad d_H(\mathcal{M}_{\text{put}}, \mathcal{M}) \leq C_* \bar{\tau},$$

measured in each frequency slice induced by M .

4. (Normal disc bundle, slice-wise and lifted) For each $\hat{z} \in \widehat{\mathcal{M}}_{\text{put}}$ and each slice k , let $E_{hi}^{(k)}(\hat{z})$ be the top $(d-p)$ -eigenspace of $\nabla^2 F_k(\hat{z}^{(k)})$. Define the slice-wise normal discs

$$\widehat{D}_{\hat{z},k}^{\text{norm}} := \{ \hat{z}^{(k)} + v^{(k)} : v^{(k)} \in E_{hi}^{(k)}(\hat{z}), \|v^{(k)}\| \leq \alpha \bar{\tau} \},$$

with a controlled $\alpha > 0$, and set $\widehat{D}_{\hat{z}}^{\text{norm}} := \times_{k=1}^c \widehat{D}_{\hat{z},k}^{\text{norm}}$. Then $\{\widehat{D}_{\hat{z}}^{\text{norm}}\}_{\hat{z} \in \widehat{\mathcal{M}}_{\text{put}}}$ forms a $C^{1,1}$ disc bundle over $\widehat{\mathcal{M}}_{\text{put}}$ in the transform domain. The lifted family

$$D_z^{\text{norm}} := M^{-1}(\widehat{D}_{Mz}^{\text{norm}}), \quad z \in \mathcal{M}_{\text{put}},$$

is a t -normal disc bundle over \mathcal{M}_{put} in \mathbb{K}_c^d .

Proof Sketch. Each local $F_{k,i}$ satisfies the Euclidean ASDf conditions of [18, Definition 16]. Since the local models agree to first order on overlaps, the partition-of-unity combination $F_k = \sum_i \phi_{k,i} F_{k,i}$ retains the same regularity and curvature control. [18, Lemma 15] ensures that the corresponding zero-gradient set forms a $C^{1,1}$ manifold with controlled reach and proximity. Stacking across slices and mapping back via the orthogonal transform M^{-1} preserves all estimates, producing \mathcal{M}_{put} and its t -normal disc bundle. \square

Finite Family of t -Normal Disc Bundles To further obtain a smooth, globally defined distance potential from locally defined quadratic models, we follow the slice-wise approach of Fefferman et al. [18] by first defining local quadratic distance models on each coordinate patch of \mathcal{M}_k and then gluing them together with a smooth partition of unity to obtain a global slice-wise potential F_k . Let $\{(\Gamma_{k,i}, \phi_{k,i})\}_{i \in I_k}$ be a $C^{1,1}$ atlas of \mathcal{M}_k , where $\Gamma_{k,i}$ is a local chart and $\{\phi_{k,i}\}$ is a smooth partition of unity subordinate to this cover. On each $\Gamma_{k,i}$, define a local model $F_{k,i}$ that approximates the squared distance to $\Gamma_{k,i}$ up to second order. Then set

$$F_k(y) := \sum_{i \in I_k} \phi_{k,i}(y) F_{k,i}(y), \quad y \in \mathcal{M}_{k,\bar{\tau}}, \quad (12)$$

where $\mathcal{M}_{k,\bar{\tau}} = \{y \in \mathbb{R}^d : \text{dist}(y, \mathcal{M}_k) < \bar{\tau}\}$ and $\bar{\tau} = c_4 \tau$ with a controlled $c_4 \in (0, 1)$. Since local models agree up to first order on overlaps and $\phi_{k,i}$ are smooth, F_k inherits the ASDf regularity properties of the Euclidean construction.

The smooth potentials F_k constructed above determine, through their gradients and Hessians, a system of local normal directions around each slice manifold \mathcal{M}_k . These directions define short, tube-like neighborhoods—local normal discs—that describe how the manifold extends in its nearby ambient space. The collection of such discs over all slices forms a smooth t -normal disc bundle, representing the local geometric envelope induced by the t -cylinder packet.

We then restrict attention to a finite, parameterized family of these disc bundles, where the parameters are inherited from the underlying t -cylinder packet (e.g., centers, orientations, and curvature bounds). This family serves as a discretized hypothesis space encompassing all admissible manifold geometries consistent with the cylinder construction. It connects the smooth analytic modeling established via the ASDf with the data-driven refinement that follows: each candidate t -manifold will later appear as a smooth section of one such disc bundle, and the estimation problem reduces to selecting the section most consistent with the observed tensor data.

In Fefferman et al. [18], this role is played by an *exhaustive family of disc bundles* \bar{D} , each built from a finite *cylinder packet*, such that every manifold $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ appears as a smooth section of at least one disc bundle in \bar{D} . We now formulate its tensorial analogue.

In the t-product setting, each frequency slice \mathcal{M}_k admits a local normal bundle constructed from the ASDF F_k (see Lemma 4), and the family $\{\hat{D}_{\hat{z}}^{\text{norm}}\}_{\hat{z} \in \hat{\mathcal{M}}_{\text{put}}}$ forms a global $C^{1,1}$ disc bundle in the transform domain. To make this representation computationally tractable, we discretize the geometric parameter space (centers, tangent orientations, and curvature scales) on controlled grids within the transform domain.

Definition 11 (Finite family of t-normal disc bundles). Let $\hat{\mathcal{L}}_\tau \subset \mathbb{R}^{c \times d}$ be a τ -lattice covering the transform domain, and let $\hat{\mathcal{G}}_{p,c}(\tau)$ be a finite $(C\tau^2)$ -net of p -dimensional subspaces in each slice of the Grassmannian. For each choice of center $\hat{\ell} \in \hat{\mathcal{L}}_\tau$, orientation $\hat{\Pi} \in \hat{\mathcal{G}}_{p,c}(\tau)$, and curvature parameter $A \in [0, C/\tau]$, define the slice-wise discs

$$\hat{D}_k^{\text{loc}}(\hat{\ell}, \hat{\Pi}, A) := \left\{ \hat{\ell}^{(k)} + v^{(k)} : v^{(k)} \in (\hat{\Pi}^{(k)})^\perp, \|v^{(k)}\| \leq \alpha\tau \right\},$$

and their product

$$\hat{D}^{\text{loc}}(\hat{\ell}, \hat{\Pi}, A) := \bigtimes_{k=1}^c \hat{D}_k^{\text{loc}}(\hat{\ell}, \hat{\Pi}, A) \subset \mathbb{R}^{c \times d}.$$

The associated ASDF $F_{\hat{\ell}, \hat{\Pi}, A}$ is obtained by inserting $(\hat{\ell}, \hat{\Pi}, A)$ into the slice-wise ASDF template of Lemma 4. The *finite t-disc family in the transform domain* is

$$\hat{\bar{D}}_t := \left\{ \hat{D}^{\text{loc}}(\hat{\ell}, \hat{\Pi}, A) : \hat{\ell} \in \hat{\mathcal{L}}_\tau, \hat{\Pi} \in \hat{\mathcal{G}}_{p,c}(\tau), A \in [0, C/\tau] \right\},$$

and the lifted family in the tensor domain is $\bar{D}_t = \{M^{-1}(\hat{D}) : \hat{D} \in \hat{\bar{D}}_t\}$.

Lemma 5 (Exhaustiveness of the t-disc family). *For every t-manifold $\mathcal{M} \in \mathcal{G}(p, V, \tau)$, there exists at least one $\hat{D}^{\text{loc}} \in \hat{\bar{D}}_t$ and a $C^{1,1}$ section $\hat{s} : \hat{\mathcal{M}}_{\text{put}} \rightarrow \hat{D}^{\text{loc}}$ such that $\hat{\mathcal{M}} = \{\hat{y} + \hat{s}(\hat{y}) : \hat{y} \in \hat{\mathcal{M}}_{\text{put}}\}$ and $\|\nabla \hat{s}\|_\infty \leq C/\tau$. Consequently, $\hat{\bar{D}}_t$ (and its lifted version \bar{D}_t) forms an exhaustive family: every admissible t-manifold is realized as a smooth section of at least one disc bundle in this finite collection.*

Proof Sketch. The argument follows the discretization scheme of Fefferman et al. [18, Lemma 17 and Corollary 18]. Since \mathcal{M} admits a finite anchor patch set (Lemma 2), its local geometry in each slice is characterized by finitely many parameters: center position, tangent orientation, and curvature scale. Quantizing these parameters on scales $(\tau, \tau^2, 1/\tau)$ yields a finite set $\hat{\bar{D}}_t$ that covers all admissible slice-wise normal configurations. Each local patch of $\hat{\mathcal{M}}_{\text{put}}$ lies within $O(\tau)$ Hausdorff distance of some element of $\hat{\bar{D}}_t$, and the corresponding global section \hat{s} is reconstructed by the slice-wise gluing procedure of Lemma 6. \square

This construction provides the geometric discretization required for the next stage of data refinement. Each disc bundle in the finite family \bar{D}_t corresponds to one admissible geometric configuration, acting as an interface representing a possible t-manifold in $\mathcal{G}(p, V, \tau)$. Recovering the underlying manifold from data thus becomes a *section-selection problem*: among all smooth sections of these bundles, the goal is to identify the one whose geometry best agrees with the observed samples.

Local Section Construction (Whitney Interpolation in the Transform Domain) The exhaustiveness result (Lemma 5) shows that every admissible t-manifold can be realized as a smooth section of some t-normal disc bundle constructed over the putative manifold \mathcal{M}_{put} . Hence, the problem of identifying the true underlying manifold from observed tensor data reduces to finding, among all possible sections, the one that best aligns with the data. Intuitively, while \mathcal{M}_{put} provides a geometrically valid template, it is not guaranteed to coincide with the true manifold; the data themselves must guide a refinement of this template by determining a locally consistent section of the associated disc bundle.

Formally, let $\{x_1, \dots, x_n\} \subset \mathbb{K}_c^d$ denote the observed data and $\hat{x}_i = Mx_i$ their transform-domain representations. Around each anchor point $z_j \in \mathcal{M}_{\text{put}}$ with its transform $\hat{z}_j = Mz_j$, we consider the neighboring transformed samples $\{\hat{x}_\ell^{(k)}\}$ on each frequency slice. Each such sample is decomposed into its tangent and normal components relative to the estimated tangent space $T_{\hat{z}_j^{(k)}} \hat{\mathcal{M}}_{\text{put}}^{(k)}$:

$$\hat{x}_\ell^{(k)} = \hat{u}_\ell^{(k)} + \hat{v}_\ell^{(k)}, \quad \hat{u}_\ell^{(k)} \in T_{\hat{z}_j^{(k)}} \hat{\mathcal{M}}_{\text{put}}^{(k)}, \quad \hat{v}_\ell^{(k)} \in (T_{\hat{z}_j^{(k)}} \hat{\mathcal{M}}_{\text{put}}^{(k)})^\perp. \quad (13)$$

Following the Whitney field interpolation framework of Fefferman et al. [18, Section 13], we fit, for each slice k , a smooth map $\hat{s}_{j,k} : \hat{u}^{(k)} \mapsto \hat{v}^{(k)}$ with bounded C^2 -norm $\|\hat{s}_{j,k}\|_{C^2} \leq C/\tau$, minimizing the local reconstruction error

$$\sum_{\ell} \|\hat{s}_{j,k}(\hat{u}_{\ell}^{(k)}) - \hat{v}_{\ell}^{(k)}\|^2.$$

If such a function exists, we call $\hat{s}_{j,k}$ an approximately optimal local section of the t-normal disc bundle near z_j . In geometric terms, $\hat{s}_{j,k}$ plays the same role as the ϵ -optimal interpolant in [18, Definition 20]: it defines a smooth local graph that passes through the region of the bundle most consistent with both the local data and the curvature constraints inherited from \mathcal{M}_{put} . The C^2 bound ensures geometric regularity comparable to that of the underlying manifold, while the optimality condition guarantees that the fitted section approximates the data as closely as any admissible smooth map of similar regularity. These slice-wise sections will then be smoothly patched together across overlaps to yield a globally consistent section representing the reconstructed t-manifold.

Patching Local t-Sections into a Global Manifold. Having obtained the slice-wise local sections $\hat{s}_{j,k} : \hat{u}^{(k)} \mapsto \hat{v}^{(k)}$ through Whitney interpolation around each anchor point $z_j \in \mathcal{M}_{\text{put}}$, we now glue these locally fitted sections into a single global smooth section that defines the final manifold \mathcal{M}_{fin} . This step parallels the global assembly in Fefferman et al. [18], adapted to the slice-wise t-product setting.

Lemma 6 (Construction of the Final t-Manifold). *Let $\mathcal{M}_{\text{put}} = M^{-1}(\widehat{\mathcal{M}}_{\text{put}})$ and let $X = \{x_i\}_{i=1}^N$ be the anchor patch set as in Lemma 2. Then there exists a global section $s : \mathcal{M}_{\text{put}} \rightarrow D^{\text{norm}}$ whose transform-domain components $\hat{s}_k : [M\mathcal{M}_{\text{put}}]_k \rightarrow [MD^{\text{norm}}]_k$ are $C^{1,1}$ for all $k \in [c]$, such that:*

1. $s(x_i) = 0$ for all $x_i \in X$;
2. $\|\nabla \hat{s}_k\|_{\infty} \leq C/\tau$ for all $k \in [c]$;
3. The resulting manifold

$$\widehat{\mathcal{M}}_{\text{fin}} := \{y + \hat{s}(y) : y \in \widehat{\mathcal{M}}_{\text{put}}\}, \quad \mathcal{M}_{\text{fin}} := M^{-1}(\widehat{\mathcal{M}}_{\text{fin}}),$$

$$\text{satisfies } \text{reach}_{\text{spec}}(\mathcal{M}_{\text{fin}}) \geq c_6\tau \text{ and } d_H(\mathcal{M}_{\text{fin}}, \mathcal{M}_{\text{put}}) \leq C\tau.$$

Proof Sketch. The argument follows the gluing framework of Fefferman et al. [18, Sections 14–15], adapted to the slice-wise structure of the t-product geometry. For each anchor point $z_j \in \mathcal{M}_{\text{put}}$ with transform $\hat{z}_j = Mz_j$, let $\hat{\Gamma}_{j,k} \subset \mathbb{R}^d$ denote a local chart of $\widehat{\mathcal{M}}_{\text{put}}^{(k)}$ centered at $\hat{z}_j^{(k)}$. The fitted function $\hat{s}_{j,k}$ defines a local $C^{1,1}$ section on $\hat{\Gamma}_{j,k} \cap B(\hat{z}_j^{(k)}, c_6\tau)$ with controlled derivative bound $\|\nabla \hat{s}_{j,k}\|_{\infty} \leq C/\tau$. By construction, these sections are consistent on overlapping neighborhoods up to first order, ensuring that tangent and normal components vary continuously across slices.

Let $\{\theta_j\}$ be a smooth partition of unity on $\widehat{\mathcal{M}}_{\text{put}}$ subordinate to the cover $\{\hat{\Gamma}_j\}$, where $\hat{\Gamma}_j := \times_k \hat{\Gamma}_{j,k}$. We form the global slice-wise section by

$$\hat{s}(y) := \sum_{j=1}^N \theta_j(y) \hat{s}_j(y), \quad \hat{s}_j := \times_{k=1}^c \hat{s}_{j,k}.$$

Since each $\hat{s}_{j,k}$ is $C^{1,1}$ and the partition $\{\theta_j\}$ is smooth, the resulting \hat{s} remains $C^{1,1}$ with curvature bounded by C/τ on each slice. Stacking the components across all frequencies yields $\hat{s} = (\hat{s}_1, \dots, \hat{s}_c)$, and applying M^{-1} gives the corresponding section $s = M^{-1}(\hat{s})$ in \mathbb{K}_c^d .

As in [18, Section 15], the manifold

$$\widehat{\mathcal{M}}_{\text{fin}} = \{y + \hat{s}(y) : y \in \widehat{\mathcal{M}}_{\text{put}}\}$$

can be viewed as a $C^{1,1}$ normal perturbation of $\widehat{\mathcal{M}}_{\text{put}}$. By the standard stability properties of the reach under small $C^{1,1}$ perturbations, one obtains $\text{reach}(\widehat{\mathcal{M}}_{\text{fin}}) \geq c_6\tau$ at the slice level. Because M^{-1} is orthogonal, these properties transfer directly to $\mathcal{M}_{\text{fin}} = M^{-1}(\widehat{\mathcal{M}}_{\text{fin}})$. We do not attempt to reproduce the precise constant tracking of [18]; the argument only relies on the structural regularity of the gluing construction and the orthogonality of M . \square

Existence of a Good Section in the Finite t-Disc Family The above constructions start from t-cylinder packets, through slice-wise ASDFs, and end with a family of t-normal disc bundles, yielding a finite and geometrically controlled representation of all admissible manifolds in $\mathcal{G}(p, V, \tau)$. Each disc bundle in this family corresponds to a local geometric configuration determined by the underlying cylinder parameters.

We denote this finite, exhaustive collection by \bar{D}_t (Definition 11). By Lemma 5, every admissible manifold $\mathcal{M} \in \mathcal{G}(p, V, \tau)$ can be represented as a smooth section of at least one local bundle $\hat{D}^{\text{loc}} \in \hat{D}_t$. This discretization reduces the original search over the manifold space $\mathcal{G}(p, V, \tau)$ to finitely many convex optimization problems, one associated with each bundle.

For any $\hat{D}^{\text{loc}} \in \hat{D}_t$, let $\mathcal{S}(\hat{D}^{\text{loc}})$ denote the convex set of admissible $C^{1,1}$ sections satisfying $\|\nabla \hat{s}\|_\infty \leq C/\tau$. We define the expected quadratic loss of a section as

$$L(\hat{s}) := \mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, M^{-1}(\text{graph}(\hat{s})))].$$

Because the map $\hat{s} \mapsto L(\hat{s})$ is convex on $\mathcal{S}(\hat{D}^{\text{loc}})$, the optimization

$$L^*(\hat{D}^{\text{loc}}) := \inf_{\hat{s} \in \mathcal{S}(\hat{D}^{\text{loc}})} L(\hat{s})$$

can be handled using convex programming techniques, as in the Euclidean setting [18, Section 13].

We declare that a manifold consistent with the data exists if there is at least one $\hat{D}^{\text{loc}} \in \hat{D}_t$ such that

$$L^*(\hat{D}^{\text{loc}}) \leq \eta,$$

where $\eta > 0$ is a prescribed tolerance. If no such bundle admits a section below this threshold, we reject the manifold hypothesis. This test constitutes the t-product analogue of the manifold existence criterion in [18].

Quadratic Loss Stability for t-Manifolds The previous step ensures that, if the manifold hypothesis is accepted, then there exists at least one bundle $\hat{D}^{\text{loc}} \in \hat{D}_t$ and an optimal section $\hat{s}^* \in \mathcal{S}(\hat{D}^{\text{loc}})$ whose graph $M^{-1}(\text{graph}(\hat{s}^*))$ provides the best-fitting t-manifold in that class. We now show that this t-manifold, denoted \mathcal{M}_{fin} , achieves a population loss within a universal constant factor of the globally optimal manifold in $\mathcal{G}(p, V, \tau)$.

Lemma 7 (Quadratic Loss Stability for t-Manifolds). *Let $\mathcal{M}_{\text{opt}} \in \mathcal{G}(p, V, \tau)$ satisfy*

$$\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{opt}})] \leq \inf_{\mathcal{M} \in \mathcal{G}(p, V, \tau)} \mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M})] + \epsilon.$$

Let $\mathcal{M}_{\text{fin}} = M^{-1}(\hat{\mathcal{M}}_{\text{fin}})$ be the manifold obtained from the optimal section \hat{s}^ of one of the bundles in \hat{D}_t . Then there exists a universal constant $C_0 > 0$ such that*

$$\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{fin}})] \leq C_0 \left(\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{opt}})] + \epsilon \right).$$

Proof Sketch. The argument parallels Fefferman et al. [18, Section 16], adapted to the slice-wise geometry of \mathbb{K}_c^d . Let $\hat{\mathcal{M}}_{\text{opt}} = M(\mathcal{M}_{\text{opt}})$ and $\hat{\mathcal{M}}_{\text{fin}} = M(\mathcal{M}_{\text{fin}})$. Each slice $\hat{\mathcal{M}}_{\text{opt}}^{(k)}, \hat{\mathcal{M}}_{\text{fin}}^{(k)} \subset \mathbb{R}^d$ satisfies the ASDF structural bounds (Lemma 4).

Following the proof of [18, Lemma 24], we decompose the data distribution as $\mathbb{P} = \mathbb{P}_{\text{in}} + \mathbb{P}_{\text{out}}$, where \mathbb{P}_{in} is supported near the anchor patches and \mathbb{P}_{out} on their complement. For \mathbb{P}_{out} , the spectral reach assumption implies that all samples lie at distance at least $c_7\tau$ from both manifolds, yielding the geometric comparison

$$\mathbb{E}_{\mathbb{P}_{\text{out}}}[\text{dist}^2([Mx]_k, \hat{\mathcal{M}}_{\text{fin}}^{(k)})] \leq (1 + c_7) \mathbb{E}_{\mathbb{P}_{\text{out}}}[\text{dist}^2([Mx]_k, \hat{\mathcal{M}}_{\text{opt}}^{(k)})].$$

For \mathbb{P}_{in} , the local section $\hat{s}_{j,k}^*$ acts as an ϵ -optimal interpolant in the sense of [18, Definitions 10]. Convexity of the squared loss and the stability of optimal interpolants imply that

$$\mathbb{E}_{\mathbb{P}_{\text{in}}}[\text{dist}^2([Mx]_k, \hat{\mathcal{M}}_{\text{fin}}^{(k)})] \leq C_0 \left(\mathbb{E}_{\mathbb{P}_{\text{in}}}[\text{dist}^2([Mx]_k, \hat{\mathcal{M}}_{\text{opt}}^{(k)})] + \epsilon \right).$$

Summing over all frequency slices and invoking Parseval's identity,

$$\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{fin}})] = \sum_{k=1}^c \mathbb{E}_{\mathbb{P}}[\text{dist}^2([Mx]_k, \hat{\mathcal{M}}_{\text{fin}}^{(k)})],$$

we obtain the desired global bound. \square

C.3 Final Proof of Theorem 1

We sketch the main argument for Theorem 1. The proof adapts the framework of Fefferman et al. [18] to the slice-wise tensor setting induced by the t-product. Our goal is to highlight the essential geometric and statistical reasoning; precise control of constants can be refined following the methodology in [18].

Step 1: Discretizing the geometric hypothesis space. By Lemma 4, local quadratic approximations of the tensor data define approximate squared distance functions (ASDFs) whose zero-gradient sets yield putative manifolds with bounded reach and curvature. Lemma 5 further constructs a finite, exhaustive family of t-normal disc bundles \bar{D}_t so that every $\mathcal{M} \in \mathcal{G}(p, V, \tau)$ is the graph of a $C^{1,1}$ section of some member of \bar{D}_t . This transforms the infinite, nonconvex manifold class into finitely many convex section spaces.

Step 2: Local-to-global reconstruction. For any bundle $D^{\text{norm}} \in \bar{D}_t$, Whitney interpolation and partition-of-unity gluing (Lemma 6) yield a global smooth section $s : \mathcal{M}_{\text{put}} \rightarrow D^{\text{norm}}$ whose graph \mathcal{M}_{fin} satisfies $\text{reach}(\mathcal{M}_{\text{fin}}) \geq c_8 \tau$ and $d_H(\mathcal{M}_{\text{fin}}, \mathcal{M}_{\text{put}}) \leq C\tau$. Hence any admissible t-manifold can be consistently reconstructed from finitely many local tensor patches.

Step 3: Convex optimization over sections. For each disc bundle $\hat{D}^{\text{loc}} \in \hat{\bar{D}}_t$, consider the convex family of admissible sections $\mathcal{S}(\hat{D}^{\text{loc}})$. Define the population and empirical quadratic losses:

$$L(\hat{s}) = \mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, M^{-1}(\text{graph}(\hat{s})))], \quad L_{\text{emp}}(\hat{s}) = \frac{1}{n} \sum_{i=1}^n \text{dist}^2(x_i, M^{-1}(\text{graph}(\hat{s}))).$$

The population minimizer within $\mathcal{S}(\hat{D}^{\text{loc}})$ is $L^*(\hat{D}^{\text{loc}}) = \inf_{\hat{s}} L(\hat{s})$. We say the manifold hypothesis holds if there exists $\hat{D}^{\text{loc}} \in \hat{\bar{D}}_t$ with $L^*(\hat{D}^{\text{loc}}) \leq \eta$, for a decision threshold $\eta > 0$ to be specified below.

Step 4: Statistical uniform convergence. Consider the function class

$$\hat{\mathcal{F}} = \left\{ \hat{x} \mapsto \text{dist}^2(\hat{x}, \widehat{\mathcal{M}}) : \widehat{\mathcal{M}} \in M(\mathcal{G}(p, V, \tau)) \right\}.$$

Since the transform M is orthogonal, both the population and empirical quadratic losses decompose slice-wise:

$$L_{\text{emp}}(\widehat{\mathcal{M}}) = \sum_{k=1}^c L_{\text{emp}}^{(k)}(\widehat{\mathcal{M}}^{(k)}), \quad \mathcal{L}(\widehat{\mathcal{M}}, \mathbb{P}) = \sum_{k=1}^c \mathcal{L}^{(k)}(\widehat{\mathcal{M}}^{(k)}, \mathbb{P}).$$

Hence it suffices to control the deviation on each frequency slice up to ϵ/c , ensuring the total deviation is bounded by ϵ .

For each slice $k \in [c]$ and $r > 0$, let $U_{\hat{\mathcal{G}}_t}^{(k)}(1/r)$ denote the smallest number of Euclidean balls of radius $\sqrt{\tau r}/16$ in \mathbb{R}^d required to cover any $\widehat{\mathcal{M}}^{(k)} \subset [M(\mathcal{G}(p, V, \tau))]_k$. By the slice-wise volume and reach bounds (cf. [18, Corollary 2]),

$$U_{\hat{\mathcal{G}}_t}^{(k)}(1/r) \leq C V \left(\frac{1}{\tau^p} + \frac{1}{(\tau r)^{p/2}} \right).$$

Applying the empirical-process bound of Fefferman et al. [18, Lemma 6] to each slice with accuracy ϵ/c and confidence δ/c yields

$$\Pr \left[\sup_{\widehat{\mathcal{M}}^{(k)}} |L_{\text{emp}}^{(k)}(\widehat{\mathcal{M}}^{(k)}) - \mathcal{L}^{(k)}(\widehat{\mathcal{M}}^{(k)}, \mathbb{P})| < \frac{\epsilon}{c} \right] \geq 1 - \frac{\delta}{c},$$

provided that the sample size satisfies

$$n \geq \frac{C c^2}{\epsilon^2} \left(U_{\hat{\mathcal{G}}_t}^{(k)}(c/\epsilon) \log^4 \frac{c U_{\hat{\mathcal{G}}_t}^{(k)}(c/\epsilon)}{\epsilon} + \log \frac{c}{\delta} \right).$$

A union bound over all c frequency slices yields

$$\Pr \left[\sup_{\widehat{\mathcal{M}} \in \mathcal{M}(\mathcal{G}(p, V, \tau))} |L_{\text{emp}}(\widehat{\mathcal{M}}) - \mathcal{L}(\widehat{\mathcal{M}}, \mathbb{P})| < \epsilon \right] \geq 1 - \delta.$$

Thus, for sufficiently large n , the empirical quadratic loss uniformly approximates the population loss across all admissible t-manifolds.

Let $\eta = c_9 \epsilon$ with a fixed constant $c_9 > 1$. Then, with probability at least $1 - \delta$:

- (i) If there exists \widehat{D}^{loc} and a section \widehat{s} such that $\mathcal{L}(\widehat{s}, \mathbb{P}) \leq (c_9 - 1)\epsilon$, then $L_{\text{emp}}(\widehat{s}) \leq \eta$ (acceptance).
- (ii) If for all sections \widehat{s} , $\mathcal{L}(\widehat{s}, \mathbb{P}) \geq (c_9 + 1)\epsilon$, then $L_{\text{emp}}(\widehat{s}) > \eta$ (rejection).

Moreover, for the empirical minimizer

$$\widehat{s}^* = \arg \min_{\widehat{s} \in \mathcal{S}(\widehat{D}^{\text{loc}})} L_{\text{emp}}(\widehat{s}),$$

the population risk satisfies

$$\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, M^{-1}(\text{graph}(\widehat{s}^*)))] \leq L_{\text{emp}}(\widehat{s}^*) + \epsilon \leq (c_9 + 1)\epsilon.$$

Step 5: Comparison with the population optimum. Let $\mathcal{M}_{\text{opt}} \in \mathcal{G}(p, V, \tau)$ minimize the population quadratic loss up to ϵ , and let \mathcal{M}_{fin} be the manifold induced by \widehat{s}^* . Then, by Lemma 7,

$$\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{fin}})] \leq C_0(\mathbb{E}_{\mathbb{P}}[\text{dist}^2(x, \mathcal{M}_{\text{opt}})] + \epsilon).$$

Hence, \mathcal{M}_{fin} achieves population loss within a controlled factor of the optimal t-manifold.

Step 6: Decision and correctness. The final statistical test enumerates all $\widehat{D}^{\text{loc}} \in \widehat{\widehat{D}}_t$, solves the convex optimization within each, and accepts the t-manifold hypothesis iff some bundle admits a section \widehat{s} with $L_{\text{emp}}(\widehat{s}) \leq \eta$. With probability at least $1 - \delta$:

(Near case) If there exists $\mathcal{M} \in \mathcal{G}(p, CV, \tau/C)$ with average squared distance $\leq C\epsilon$, then an admissible section achieves $L_{\text{emp}} \leq \eta$ and the test accepts.

(Far case) If all $\mathcal{M} \in \mathcal{G}(p, V/C, C\tau)$ satisfy average squared distance $> \epsilon/C$, then every empirical loss exceeds η and the test rejects.

Thus, for sufficiently large n , the test distinguishes the near and far regimes with confidence $1 - \delta$, completing the proof of Theorem 1.

Remark 6 (On the Computational Cost of Cylinder Packet Enumeration). While each local optimization step involved in constructing candidate t-manifolds, such as fitting local charts or evaluating empirical loss, can be carried out in polynomial time with respect to the sample size n and the ambient parameters c and d , the overall computational cost of the testing procedure is dominated by the enumeration of admissible *t-cylinder packets*.

Each t-cylinder packet represents a discrete configuration of anchor patches and tangent orientations across frequency slices, and every candidate t-manifold is obtained by gluing local charts defined by one such packet. The number of admissible configurations grows combinatorially with the geometric parameters governing the manifold class $\mathcal{G}(p, V, \tau)$, making the enumeration phase inherently non-polynomial in complexity.

This reflects a fundamental limitation of nonparametric geometric model selection: testing manifold hypotheses without restrictive parametric assumptions inevitably encounters combinatorial explosion in geometric configurations. The proposed method inherits and extends the general strategy of Fefferman et al. [18] to the t-product setting, including the use of chart-based approximation, geometric regularity control, and statistical estimation. When $c = 1$, the construction naturally reduces to the Euclidean case in their framework.

D Theoretical Analysis For t-Manifold Fitting

We now turn to the *second aspect* of **Q2**, which concerns the recovery of the underlying t-manifold structure from finite, potentially noisy samples. While Theorem 1 establishes that the data may lie near a smooth low-dimensional t-manifold, the next step is to construct an explicit estimator $\widehat{\mathcal{M}}$ that approximates the true manifold \mathcal{M} with quantifiable geometric accuracy.

Given noisy samples in \mathbb{K}_c^d on or near a smooth low-dimensional t-manifold, can we estimate the manifold from data?

We address this fitting problem by extending the geometric–statistical framework of Yao et al. [74] from Euclidean manifold fitting to the t-product setting. The extension follows the same conceptual design: *local smoothing through a contraction mapping, followed by global reconstruction via level-set alignment*. All constructions are extended from [74] to be compatible with the algebraic structure of \mathbb{K}_c^d , where addition, averaging, and projection are performed using the t-product and \mathbb{K}_c -linear maps. The proof of Theorem 2 proceeds in two similar stages to [74]:

1. *Local smoothing via contraction mappings.* Construct a smooth contraction map $G : \mathbb{K}_c^d \rightarrow \mathbb{K}_c^d$ in two steps: first, compute a preliminary estimate $F(y)$ using *isotropic* weights to determine the contraction direction; second, define the refined projection $G(y)$ as a weighted average over an *anisotropic* “straw-like” neighborhood aligned with that direction. All coefficients are defined slice-wise in the frequency domain adopting the arguments in [74] to ensure $*$ -compatibility.
2. *Global reconstruction via level sets.* From the preliminary contraction field $F(y)$, we define a constraint on contraction directions through a fixed projection operator Π^* , and extract a reference submanifold $\widetilde{\mathcal{M}} := \{y : \Pi^* * (F(y) - y) = 0\}$ as a smooth zero-level set in \mathbb{K}_c^d . Applying the refined contraction map G to this intermediate t-manifold yields the final estimator $\widehat{\mathcal{M}} = G(\widetilde{\mathcal{M}})$.

Each step mirrors its Euclidean counterpart in [74] while remaining algebraically faithful to the t-product. The result provides a data-driven reconstruction of t-manifolds that preserves both smoothness and structural consistency across frequency components. We also provide a preliminary numerical evaluation of the proposed t-manifold fitting method on synthetic data in Appendix D.2, demonstrating its ability to approximately recover the underlying t-manifold structures under noisy observations.

D.1 Proof Strategy

The proof proceeds by constructing a preliminary smooth t-vector field $F(y)$ on a local neighborhood Γ of \mathcal{M} to estimate the contraction direction, and defining a refined projection $G(y)$ via anisotropic smoothing aligned with this direction. We then define a candidate manifold $\widetilde{\mathcal{M}}$ as the level set of the normal component of the drift $F(y) - y$, determined by a fixed projection t-matrix Π^* . The final estimator $\widehat{\mathcal{M}} = G(\widetilde{\mathcal{M}})$ inherits the C^2 regularity of the construction and remains close to \mathcal{M} in Hausdorff distance with high probability.

Step I: Local Projection via Weighted Contraction Motivated by [74], the first step in recovering a t-manifold from noisy data is to design a pointwise correction mechanism that pushes each sample $y \in \mathbb{K}_c^d$ closer to the underlying clean t-manifold \mathcal{M} . This requires careful consideration, since noise may distort both the location and orientation of the sample, and the manifold structure is not directly observable. Nevertheless, if we assume that y lies in a small neighborhood around \mathcal{M} , we can hope to exploit the surrounding data geometry to correct its position. The guiding idea is that the clean manifold induces a directional asymmetry in the neighborhood of y —samples in directions orthogonal to the manifold fluctuate more severely than those tangent to it in the frequency domain slice-wisely [74]. By detecting and leveraging this asymmetry, we define a smooth contraction map $G : \mathbb{K}_c^d \rightarrow \mathbb{K}_c^d$ that approximates a projection onto \mathcal{M} .

To describe this construction in the t-product setting, we must be careful to preserve the algebraic structure of \mathbb{K}_c^d . While the definitions ultimately rely on proximity in each frequency slice $[My]_k \in \mathbb{R}^d$, we express the construction in terms of operations in the original domain \mathbb{K}_c^d .

We begin by defining a neighborhood $\Gamma \subset \mathbb{K}_c^d$ around the manifold:

$$\Gamma := \{y \in \mathbb{K}_c^d : \exists z \in \mathcal{M}, \text{ such that } \|[My]_k - [Mz]_k\|_2 \leq C\sigma, \quad \forall k = 1, \dots, c\}.$$

This set contains all points whose spectral components are each within radius $C\sigma$ of the manifold in the corresponding frequency slice. The constant C can be chosen based on the manifold curvature and Gaussian tail bounds.

Now fix $y \in \Gamma$. The construction of the contraction map proceeds in two phases: first estimating the contraction direction, and then refining the position.

Phase 1: Direction Estimation. We define a smooth local average $F(y) \in \mathbb{K}_c^d$ using isotropic weights. This vector is constructed as a t-product weighted sum:

$$F(y) := \sum_{i=1}^n \alpha_i(y) * x_i, \quad (14)$$

where the isotropic weights $\alpha_i(y) \in \mathbb{K}_c$ are defined slice-wise in the frequency domain. For each $k = 1, \dots, c$, we follow [74, Eq. (3.4)]:

$$[M\alpha_i(y)]_k := \frac{\left(1 - \frac{\|[My]_k - [Mx_i]_k\|_2^2}{r_0^2}\right)_+^\gamma}{\sum_{j=1}^n \left(1 - \frac{\|[My]_k - [Mx_j]_k\|_2^2}{r_0^2}\right)_+^\gamma}, \quad (15)$$

where $r_0 = C_0\sigma$ is the smoothing radius and $\gamma \geq 2$ ensures smoothness. By construction, $\sum_{i=1}^n [M\alpha_i(y)]_k = 1$ for every k , hence $F(y)$ is a slice-wise average in the frequency domain and is $*$ -compatible.

Let $v := F(y) - y$. We define the slice-wise projection t-matrix \tilde{U} onto the estimated contraction direction:

$$[M\tilde{U}]_k := \frac{[Mv]_k [Mv]_k^\top}{\|[Mv]_k\|_2^2}. \quad (16)$$

Lemma 8 (Direction Estimation Error). *Let $F(y)$ be the empirical direction estimator defined in Step I (Phase 1), and let \tilde{U} be the associated projection operator. Under the sample complexity $n = \mathcal{O}(\sigma^{-(p+3)})$, for any $y \in \Gamma$, the estimated normal direction satisfies:*

$$\|\tilde{U} - U^*\|_{\mathbb{K}_c^d} \leq C\sigma\sqrt{\log(1/\sigma)}$$

with high probability.

Proof Sketch. This result is the t-product extension of [74, Theorem 3.5]. For each frequency slice k , the weights $\alpha_i(y)$ are constructed using the isotropic kernel as defined in [74, Eq. (3.4)]. Applying [74, Theorem 3.5] slice-wisely implies that for each k , we have $\|[M\tilde{U}]_k - [MU^*]_k\|_F \leq C'\sigma\sqrt{\log(1/\sigma)}$ with high probability. By a union bound, the slice-wise error bounds hold simultaneously for all $k = 1, \dots, c$ with high probability. On this event, summing the squared errors over c slices yields the global bound in \mathbb{K}_c^d . \square

Phase 2: Position Refinement. To achieve optimal geometric accuracy while maintaining smoothness, we construct the refined estimator $G(y)$ using an *anisotropic* “straw-like” neighborhood aligned with \tilde{U} . For each sample x_i , we decompose its displacement relative to y into normal and tangent components in the frequency domain by following [74, Eq. (4.13)]:

$$u_{ik} := [M\tilde{U}]_k([Mx_i]_k - [My]_k), \quad v_{ik} := ([Mx_i]_k - [My]_k) - u_{ik}. \quad (17)$$

We then define refinement weights $\beta_i(y) \in \mathbb{K}_c$ using distinct bandwidths for these components in the frequency domain slice-wisely: a larger radius $r_{\text{ctr}} = C_1\sigma\sqrt{\log(1/\sigma)}$ along the estimated *contraction direction*, and a smaller radius $r_{\text{orth}} = C_2\sigma$ along the *remaining orthogonal directions*.

Specifically, the weights are defined slice-wise for each $k = 1, \dots, c$ as:

$$[M\beta_i(y)]_k := \frac{w_u(\|u_{ik}\|_2) \cdot w_v(\|v_{ik}\|_2)}{\sum_{j=1}^n w_u(\|u_{jk}\|_2) \cdot w_v(\|v_{jk}\|_2)}, \quad (18)$$

where w_u and w_v follow the piecewise C^2 construction in [74, Eq. (4.15)]:

$$w_u(t) = \begin{cases} 1, & t \leq \frac{r_{\text{ctr}}}{2} \\ \left(1 - \left(\frac{2t - r_{\text{ctr}}}{r_{\text{ctr}}}\right)^2\right)^\gamma, & \frac{r_{\text{ctr}}}{2} < t \leq r_{\text{ctr}} \\ 0, & \text{otherwise} \end{cases}, \quad w_v(t) = \begin{cases} \left(1 - \frac{t^2}{r_{\text{orth}}^2}\right)^\gamma, & t \leq r_{\text{orth}} \\ 0, & \text{otherwise} \end{cases}, \quad (19)$$

with $\gamma \geq 2$ being a fixed integer. Similarly, the slice-wise normalization implies $\sum_{i=1}^n [M\beta_i(y)]_k = 1$ for every k , making $G(y)$ a valid slice-wise weighted average.

Finally, the corrected output is defined as the empirical weighted average:

$$G(y) := \sum_{i=1}^n \beta_i(y) * x_i. \quad (20)$$

The next lemma shows that this contraction map satisfies the expected error bound for all $y \in \Gamma$, enabling global manifold estimation.

Lemma 9 (Refined Contraction Accuracy). *Let $G(y)$ be the anisotropic empirical estimator defined in Step I (Phase 2), constructed using the estimated direction \tilde{U} . Let y^* be the true projection of y onto \mathcal{M} . Then, with high probability:*

$$\|G(y) - y^*\|_{\mathbb{K}_c^d} \leq C\sigma^2 \log(1/\sigma).$$

Proof Sketch. Since $G(y)$ is a t-product weighted sum, its spectral components satisfy $[MG(y)]_k = \sum_i [M\beta_i(y)]_k [Mx_i]_k$. By construction, each slice satisfies the conditions of [74, Theorem 4.5], where the anisotropic kernels w_u and w_v control the geometric bias and stochastic variance to $\mathcal{O}(\sigma^2 \log(1/\sigma))$ with probability at least $1 - \delta_k$. By applying a union bound over all $k = 1, \dots, c$ spectral slices, the conditions hold simultaneously for all components with probability at least $1 - \sum \delta_k$. As M is an unitary transform, the estimation error is bounded by the sum of slice-wise squared errors $\|G(y) - y^*\|_{\mathbb{K}_c^d}^2 = \sum_{k=1}^c \|[MG(y)]_k - [My^*]_k\|_2^2 \leq C'(\sigma^2 \log(1/\sigma))^2$, which simplifies to the desired bound. \square

Step II: Global Manifold Construction via Level Set Embedding Having defined a smooth contraction map $G : \mathbb{K}_c^d \rightarrow \mathbb{K}_c^d$, we now seek to construct a global estimator $\widehat{\mathcal{M}}$. Inspired by [74, Section 5], the key idea is to embed a smooth reference t-manifold $\widetilde{\mathcal{M}}$ inside the neighborhood Γ , and apply the map G to obtain $\widehat{\mathcal{M}} := G(\widetilde{\mathcal{M}})$.

We construct the reference t-manifold by extending the level set formulation in [74, Proposition 5.5] to the t-product case. For any $y \in \Gamma$, the t-vector field $v = F(y) - y$ approximates the local contraction direction in the frequency domain slice-wisely. We define a constraint on these directions through a fixed projection t-matrix $\Pi^* \in \mathbb{K}_c^{d \times d}$ satisfying

$$\text{rank}([M\Pi^*]_k) = d - p, \quad \forall k = 1, \dots, c. \quad (21)$$

The reference t-manifold $\widetilde{\mathcal{M}}$ is then defined as the zero set:

$$\widetilde{\mathcal{M}} := \{y \in \Gamma : \Pi^* * (F(y) - y) = 0\}. \quad (22)$$

Since $F(y)$ is C^2 smooth (inherited from the weights α_i) slice-wisely in the frequency domain, this constraint defines a C^2 level set in \mathbb{K}_c^d . As Π^* has constant rank due to its construction and the Jacobian is full rank slice-wisely according to [74, Proposition 5.5], the reference t-manifold $\widetilde{\mathcal{M}}$ is a p -dimensional (sub-)t-manifold in Γ .

We now define the final estimator as:

$$\widehat{\mathcal{M}} := G(\widetilde{\mathcal{M}}). \quad (23)$$

Since G is C^2 and locally bi-Lipschitz in Γ (as established by the smoothness of weights β_i) slice-wisely in the frequency domain, and $\widetilde{\mathcal{M}}$ is a C^2 t-manifold, it follows that $\widehat{\mathcal{M}}$ is also a p -dimensional C^2 t-manifold.

The geometric validity of this estimator is guaranteed by the following lemmas.

Lemma 10 (Global Geometry of the Reconstructed Manifold). *Let $\widetilde{\mathcal{M}}$ be the level set defined in Eq. (22), and let $\widehat{\mathcal{M}} := G(\widetilde{\mathcal{M}})$ be the final estimator. Then, with high probability:*

1. *Regularity: $\widehat{\mathcal{M}}$ is a p -dimensional C^2 t-manifold.*
2. *Hausdorff Distance: $d_H(\mathcal{M}, \widehat{\mathcal{M}}) \leq C\sigma^2 \log(1/\sigma)$.*

Proof Sketch. The proof relies on the decoupling of t-product operations in the frequency domain. First, for *Regularity*, since the kernels used in the construction of G are C^2 and the mapping is defined slice-wise, G acts as a C^2 local diffeomorphism within each frequency subspace. By [74, Proposition 5.5], the zero-level set conditions are satisfied slice-wise, ensuring that $\widetilde{\mathcal{M}}$ is a p -dimensional t-manifold where each frequency component $[\widetilde{\mathcal{M}}]_k$ is a standard p -dimensional submanifold in \mathbb{R}^d .

For the *Hausdorff Distance*, we utilize the fact that the \mathbb{K}_c^d -norm is equivalent to the ℓ_2 -sum of spectral errors (Parseval’s identity). Since Lemma 9 provides a uniform error bound of $\mathcal{O}(\sigma^2 \log(1/\sigma))$ for each spectral component $[MG(y)]_k$ relative to $[My^*]_k$, the global distance in \mathbb{K}_c^d is bounded by the same order. Applying [74, Theorem 5.4] and a union bound over the c spectral slices, we conclude that $\widehat{\mathcal{M}} = G(\widetilde{\mathcal{M}})$ remains within $\mathcal{O}(\sigma^2 \log(1/\sigma))$ of \mathcal{M} in the total t-vector space with high probability. \square

With the local and global geometric properties established in the lemmas above, we are now positioned to conclude the proof of the main theorem.

Proof of Theorem 2. The theorem follows directly from the construction in Steps I–II and the bounds established in the supporting lemmas. Specifically, Lemma 9 guarantees the pointwise convergence rate of $\mathcal{O}(\sigma^2 \log(1/\sigma))$, which implies the Hausdorff distance bound in Lemma 10. Furthermore, Lemma 10 confirms that $\widehat{\mathcal{M}}$ is a C^2 t-manifold. This verifies all claims in Theorem 2. \square

D.2 Synthetic Validation of t-Manifold Fitting

To preliminarily verify the effectiveness of the proposed t-manifold fitting method, we follow the experimental strategy of Yao et al. [74] and conduct a synthetic experiment under the t-product geometry. The objective is to evaluate the denoising accuracy of our estimator on two illustrative examples of t-manifolds: the t-circle and the t-torus. Each manifold is designed so that every frequency component corresponds to a smooth low-dimensional geometric structure—specifically, a circle for the t-circle and a torus for the t-torus—thereby enabling a clear assessment of how well the method preserves and reconstructs geometric consistency across channels.

In each experiment, we begin by constructing a smooth t-manifold consisting of $c = 5$ frequency channels, where the transform matrix M is set to the DCT. Each channel represents a distinct frequency component but shares a common underlying manifold structure in the spatial domain, ensuring geometric coherence across frequencies. The manifolds are defined as follows:

- **t-circle:** each frequency component corresponds to a one-dimensional circular manifold embedded in \mathbb{R}^2 . Each point on the manifold is parameterized by an angular variable $\theta \in [0, 2\pi)$, forming coordinates $(\cos \theta, \sin \theta)$. The intrinsic dimension is $p = 1$, and the ambient dimension is $d = 2$.
- **t-torus:** each frequency component corresponds to a two-dimensional toroidal manifold embedded in \mathbb{R}^3 . Each point on the manifold is parameterized by two angular variables $(\theta_1, \theta_2) \in [0, 2\pi)^2$, forming coordinates

$$((R + r \cos \theta_1) \cos \theta_2, (R + r \cos \theta_1) \sin \theta_2, r \sin \theta_1),$$

where R and r denote the major and minor radii, respectively. The intrinsic dimension is $p = 2$, and the ambient dimension is $d = 3$.

After defining the manifold geometry, we uniformly sample n points from each channel of the manifold to obtain n clean t-vectors, each of size $d \times 1 \times c$. These t-vectors collectively form the clean dataset representing the target t-manifold. To simulate real-world measurement perturbations,

we add Gaussian noise independently to every spatial entry across all frequency channels. The resulting noisy t-vectors serve as the observed data for manifold fitting and denoising evaluation.

For evaluation, we compute the pointwise Euclidean distances between the recovered and ground-truth manifolds, and report the average distance (Avg Dist) and maximum distance (Max Dist) aggregated over all frequency channels and multiple trials. The experiments are repeated under increasing sample sizes $n = 50, 100, 200, 400, 800$. The results are summarized in Tables D.1 and D.2.

Table D.1: Estimation accuracy on the t-circle manifold under different sample sizes.

Sample Size	Avg Dist (Mean \pm Std)	Max Dist (Mean \pm Std)
50	0.0364 ± 0.0044	0.1377 ± 0.0316
100	0.0212 ± 0.0013	0.0775 ± 0.0131
200	0.0151 ± 0.0014	0.0569 ± 0.0069
400	0.0116 ± 0.0006	0.0473 ± 0.0079
800	0.0096 ± 0.0004	0.0383 ± 0.0053

Table D.2: Estimation accuracy on the t-torus manifold under different sample sizes.

Sample Size	Avg Dist (Mean \pm Std)	Max Dist (Mean \pm Std)
50	0.1111 ± 0.0072	0.3102 ± 0.0065
100	0.0711 ± 0.0037	0.2919 ± 0.0196
200	0.0414 ± 0.0025	0.2077 ± 0.0230
400	0.0258 ± 0.0009	0.1210 ± 0.0124
800	0.0187 ± 0.0005	0.0893 ± 0.0073

The results in Tables D.1 and D.2 show a consistent decrease in both average and maximum estimation errors as the sample size increases, confirming that the proposed method converges toward the true manifold structure. These findings support the theoretical guarantees of our framework and demonstrate that the t-product geometry enables stable and accurate manifold recovery in multi-channel tensor settings. Future work will extend these experiments to higher-dimensional and application-driven t-manifolds.

E Theoretical Analysis For Learning on Approximate t-Manifolds

We now turn to *the third aspect* of **Q2**, which concerns the learning of functions defined on low-dimensional t-manifolds. While previous sections have established the existence of such manifolds near observed tensor data and provided procedures for estimating their geometry, we now ask whether meaningful functions supported on or near these structures can be efficiently learned from finite samples.

Given samples near a smooth low-dimensional t-manifold, can we learn functions that depend on the manifold structure using tensor-based models?

We investigate this question in the setting of nonparametric regression, where the unknown target function f_0 maps elements of a p -dimensional t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$ into the t-scalar space \mathbb{K}_c . The goal is to approximate f_0 using tensor neural networks (TNN) [49], which are designed to respect both the spectral geometry of \mathcal{M} and the algebraic structure of the t-product. Our analysis builds on the approximation and generalization techniques developed in recent manifold learning theory, while introducing new tools to accommodate the transform-based structure of \mathbb{K}_c .

Theoretical Setup and Main Theorem We now present a proof of Theorem 3, which establishes nonparametric learning guarantees for TNNs trained on data sampled from a low-dimensional t-manifold. Specifically, we consider the regression setting where the input data $\{x_i\}_{i=1}^n$ are independently drawn from a p -dimensional C^β t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$, and the output responses y_i are generated according to the model

$$y_i = f_0(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I_c),$$

for some unknown \mathbb{K}_c -smooth function $f_0 : \mathcal{M} \rightarrow \mathbb{K}_c$ and additive isotropic Gaussian noise. Our goal is to learn f_0 from data using a TNN predictor \hat{f}_n , trained by empirical risk minimization over a hypothesis class \mathcal{F}_n that respects the t-product algebra and the spectral geometry of \mathcal{M} .

Remark 7 (Smoothness Assumption and Structure Sheaf). The smoothness condition (A2) in Assumption 1 assumed for the target function $f_0 : \mathcal{M} \rightarrow \mathbb{K}_c$ is consistent with the structure sheaf $\mathcal{O}_{\mathcal{M}}$ of the t-manifold \mathcal{M} (see Definition 3). Specifically, the requirement that each slice $[M(f_0 \circ \varphi_\alpha^{-1})]_k$ is C^β -smooth corresponds to the sheaf-theoretic formulation of local regularity under the transform domain. This choice respects the algebraic structure induced by the t-product and allows learning theory to be developed within a frequency-wise smooth calculus framework. Alternative formulations are possible, such as defining f_0 as a scalar-valued function on \mathcal{M} or considering smoothness in the ambient Euclidean space $\mathbb{R}^{c \times d}$. However, since our goal is to preserve the tensor-algebraic coherence of the t-product, we adopt the definition given in Assumption (A2) as a natural model for function classes on t-manifolds. Exploring broader formulations beyond the structure sheaf $\mathcal{O}_{\mathcal{M}}$ is an interesting direction for future work.

We aim to prove that, under Assumption (A2) on the smoothness of f_0 , the geometry of the t-manifold \mathcal{M} , and the structure of the TNN architecture (including width, depth, and activation regularity), the estimator \hat{f}_n achieves the following convergence rate:

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

where $p_{\text{eff}} = \mathcal{O}(p \log(cd))$ denotes the effective intrinsic complexity of the learning problem, reflecting both the manifold dimension p and the channel-ambient interaction due to the t-product.

In the proof of Theorem 3, we adopt a slightly stronger and more realistic setting than the theorem statement suggests. Instead of assuming the input data $\{x_i\}_{i=1}^n$ lie exactly on a compact p -dimensional t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$, we allow them to be sampled from a ρ -neighborhood of \mathcal{M} , denoted $\mathcal{M}_\rho := \{x \in \mathbb{K}_c^d : \inf_{z \in \mathcal{M}} \|x - z\|_{\mathbb{K}_c^d} \leq \rho\}$. This generalization accommodates potential data noise or geometric perturbations, and plays an essential role in achieving a robust nonparametric approximation result.

We formally state this stronger version as follows:

Theorem 4 (Generalization bound on approximate t-manifolds). *Let $f_0 : \mathbb{K}_c^d \rightarrow \mathbb{K}_c$ be a C^β \mathbb{K}_c -smooth function supported on the ρ -neighborhood \mathcal{M}_ρ of a compact p -dimensional t-manifold*

$\mathcal{M} \subset \mathbb{K}_c^d$, with spectral reach at least $\tau > 0$ and spectral volume at most V . Suppose the sample pairs (x_i, y_i) follow the noisy model

$$y_i = f_0(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I_c), \quad x_i \in \mathcal{M}_\rho,$$

where x_i are drawn i.i.d. from a distribution ν supported on \mathcal{M}_ρ .

Then there exists a class \mathcal{F}_n of tensor neural networks as defined in Eq. (1), with architecture width and depth denoted \mathcal{W} and \mathcal{D} , such that the empirical risk minimizer $\hat{f}_n \in \mathcal{F}_n$ in Eq. (3) satisfies

$$\mathbb{E} \left[\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \right] \leq C \cdot n^{-\frac{2\beta}{2\beta + p_{\text{eff}}}}, \quad \text{where } p_{\text{eff}} := p \log(cd),$$

provided the neighborhood radius ρ is sufficiently small in terms of \mathcal{W} and \mathcal{D} , and p_{eff} . The constant C depends on $(\beta, B_0, \sigma, \tau, V, c, \log n)$, and grows polynomially in the problem parameters $(p, B_0, c, d, \log n)$.

This refined result subsumes the exact manifold case as a special situation when $\rho = 0$. Hereafter, we provide a proof of this stronger theorem, which will in turn prove Theorem 3 as a corollary.

The rest of this section follows the four-step approach introduced by Jiao et al. [29], but extends it to the t-product setting by adaptations to \mathbb{K}_c -valued regression on t-manifolds:

1. *Error decomposition:* Decompose the total error into a stochastic term and an approximation term.
2. *Approximation construction:* Construct tensor neural networks using projection and extension.
3. *Stochastic error control:* Use covering numbers and pseudo-dimension bounds for tensor neural networks.
4. *Bias-variance optimization:* Balance approximation and stochastic errors by selecting the network width and depth parameters optimally.

E.1 Learning with Approximate t-Manifold Support

The remainder of this section is devoted to proving Theorem 4. Theorem 3 then follows as a direct corollary by setting $\rho = 0$. We adopt a four-part structure modeled on the nonparametric framework of Jiao et al. [29], adapted to the t-product case.

Step 1: Error Decomposition We begin by decomposing the expected excess risk into an approximation term and a stochastic term, which allows us to control separately the expressiveness of the hypothesis class and the estimation error arising from finite samples. This result corresponds to Lemma 3.1 of Jiao et al. [29], adapted to the \mathbb{K}_c^d setting under the t-product structure. For completeness and self-containment, we include a detailed proof below.

Lemma 11 (Decomposition of Expected Excess Risk). *Let \mathcal{F}_n be a hypothesis class of functions from \mathbb{K}_c^d to \mathbb{K}_c . Suppose the training data $S := \{(x_i, y_i)\}_{i=1}^n$ are drawn i.i.d. from the model $y = f_0(x) + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_c)$. Let $\hat{f}_n \in \arg \min_{f \in \mathcal{F}_n} L_n(f)$ be the empirical risk minimizer under squared \mathbb{K}_c -norm loss. Then the expected excess risk (under squared L^2 error) satisfies*

$$\mathbb{E}_S \|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \leq \underbrace{\mathbb{E}_S [L(f_0) - 2L_n(\hat{f}_n) + L(\hat{f}_n)]}_{\text{stochastic error}} + 2 \underbrace{\inf_{f \in \mathcal{F}_n} \|f - f_0\|_{L^2(\nu)}^2}_{\text{approximation error}}. \quad (24)$$

Proof. The expected L^2 error can be written using the definition of population risk:

$$\mathbb{E}_S \|\hat{f}_n - f_0\|_{L^2(\nu)}^2 = \mathbb{E}_S [L(\hat{f}_n) - L(f_0)],$$

since the noise is zero-mean and independent of x .

Now let $f^* \in \mathcal{F}_n$ be any candidate function, in particular one that is closest to f_0 in $L^2(\nu)$ norm. Because \hat{f}_n minimizes the empirical risk over \mathcal{F}_n , we have

$$L_n(\hat{f}_n) \leq L_n(f^*),$$

and so

$$L_n(\hat{f}_n) - L(f_0) \leq L_n(f^*) - L(f_0).$$

Taking expectations on both sides yields

$$\mathbb{E}_S[L_n(\hat{f}_n) - L(f_0)] \leq L(f^*) - L(f_0),$$

since $\mathbb{E}_S[L_n(f^*)] = L(f^*)$.

Multiply both sides by 2 and add the identity

$$\mathbb{E}_S[L(\hat{f}_n) - L(f_0)] = \mathbb{E}_S\|\hat{f}_n - f_0\|_{L^2(\nu)}^2$$

to get

$$\mathbb{E}_S\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \leq \mathbb{E}_S[L(f_0) - 2L_n(\hat{f}_n) + L(\hat{f}_n)] + 2\|f^* - f_0\|_{L^2(\nu)}^2.$$

Since f^* was arbitrary in \mathcal{F}_n , we take the infimum over \mathcal{F}_n to conclude the proof. \square

Step 2: Approximation over Approximate t-Manifolds We now demonstrate that for any C^β -smooth t-function defined on a compact t-manifold, there exists a tensor neural network that approximates the function uniformly over a small ρ -neighborhood \mathcal{M}_ρ with controlled error and architecture size.

Lemma 12 (Approximation of C^β t-functions on approximate t-manifolds). *Let $f_0 : \mathbb{K}_c^d \rightarrow \mathbb{K}_c$ be a C^β - \mathbb{K}_c -smooth function defined on a compact p -dimensional t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$ with spectral reach at least $\tau > 0$ and spectral volume at most V . Let $\mathcal{M}_\rho := \{x \in \mathbb{K}_c^d : \inf_{z \in \mathcal{M}} \|x - z\|_{\mathbb{K}_c^d} \leq \rho\}$ be its ρ -neighborhood. Then for any $\delta \in (0, 1)$, there exist integers $W_0, L_0 \in \mathbb{N}^+$ and a tensor neural network $\hat{f}_n : \mathbb{K}_c^d \rightarrow \mathbb{K}_c$ with t-product structure, t-scalar width $\mathcal{W} = \mathcal{O}(d_\delta^{\beta+1} W_0 \log W_0)$ and depth $\mathcal{D} = \mathcal{O}(L_0 \log L_0)$, such that if ρ satisfies $\rho \leq \mathcal{O}\left((W_0 L_0)^{-2\beta/d_\delta} (\beta + 1)^2 d_\delta^{3\beta/2} (\sqrt{cd/d_\delta} + 1 - \delta)^{-1} (1 - \delta)^{1-\beta}\right)$, then the approximation error satisfies*

$$\|f_0(x) - \hat{f}_n(x)\|_{\mathbb{K}_c}^2 \leq C(W_0 L_0)^{-4\beta/d_\delta}, \quad \forall x \in \mathcal{M}_\rho, \quad (25)$$

where $d_\delta = \mathcal{O}(p \log(cd)/\delta^2)$.

Proof. We prove the approximation result by reducing the problem from the t-manifold $\mathcal{M} \subset \mathbb{K}_c^d$ to Euclidean slices and applying spectral projection and ReLU approximation.

Step 1): Frequency slice representation. Let $M \in \mathbb{R}^{c \times c}$ be the orthogonal transform defining the t-product. Each point $x \in \mathbb{K}_c^d$ can be mapped into the transform domain as $z_k := [Mx]_k \in \mathbb{R}^d$, for $k = 1, \dots, c$.

Then the t-manifold \mathcal{M} maps to frequency slice manifolds $\mathcal{M}_k := \{z_k : x \in \mathcal{M}\} \subset \mathbb{R}^d$, each C^β -smooth with reach $\geq \tau$, dimension p , and bounded diameter.

The target function $f_0 : \mathbb{K}_c^d \rightarrow \mathbb{K}_c$ is decomposed in frequency domain as $f_{0,k}(z_k) := [Mf_0(x)]_k$.

Step 2): Projection and extension. For each k , by [3, Theorem 3.1], there exists a linear projection $A_k \in \mathbb{R}^{d_\delta \times d}$ such that:

$$(1 - \delta)\|u - v\|_2 \leq \|A_k u - A_k v\|_2 \leq (1 + \delta)\|u - v\|_2, \quad \forall u, v \in \mathcal{M}_k, \quad (26)$$

and $A_k(\mathcal{M}_k)$ is embedded into $[-\sqrt{d/d_\delta}, \sqrt{d/d_\delta}]^{d_\delta} =: E_k$.

We define the reduced function $\tilde{f}_{0,k}(z) := f_{0,k}(A_k^{-1}z)$ on $A_k(\mathcal{M}_k)$. Since $f_{0,k}$ is C^β -smooth, $\tilde{f}_{0,k} \in \mathcal{H}^\beta(A_k(\mathcal{M}_k), B_0/(1 - \delta)^\beta)$. By Whitney's extension theorem [15], there exists an extension $\tilde{F}_{0,k}$ to E_k with the same regularity.

Step 3): Neural approximation. By [29, Theorem 3.3], for any $W_0, L_0 \geq 1$, there exists a ReLU network $\tilde{f}_{n,k} : \mathbb{R}^{d_\delta} \rightarrow \mathbb{R}$ with width $\mathcal{W} = \mathcal{O}(d_\delta^{\beta+1} W_0 \log W_0)$ and depth $\mathcal{D} = \mathcal{O}(L_0 \log L_0)$ such that

$$\sup_{z \in E_k} |\tilde{f}_{n,k}(z) - \tilde{F}_{0,k}(z)| \leq \varepsilon := CB_0(\beta + 1)^2 d_\delta^{3\beta/2} (W_0 L_0)^{-2\beta/d_\delta}. \quad (27)$$

Pulling back to the slice domain, define $f_{n,k}(z_k) := \tilde{f}_{n,k}(A_k z_k)$.

Step 4): Reconstruct global approximation. Define the approximator in the transform domain:

$$[M\hat{f}_n(x)]_k := f_{n,k}([Mx]_k) = \tilde{f}_{n,k}(A_k[Mx]_k),$$

and reconstruct in the original domain:

$$\hat{f}_n(x) := M^{-1} \left([M\hat{f}_n(x)]_1, \dots, [M\hat{f}_n(x)]_c \right).$$

In the above construction, each slice-wise approximation function $\tilde{f}_{n,k}$ may initially correspond to a ReLU network with varying width and depth, depending on the geometry and complexity of the k -th slice \mathcal{M}_k . To ensure that the global approximator \hat{f}_n can be implemented as a coherent TNN with consistent architecture across slices, we standardize the network structure by aligning all $\tilde{f}_{n,k}$ to a common depth \mathcal{D}_{\max} and width \mathcal{W}_{\max} . This alignment is achieved by padding each slice-wise network to a shared depth and width through identity layers and dummy neurons, following standard techniques in the ReLU network literature [77]. Such operations preserve the output function on the support domain while enabling consistent layer-wise stacking and spectral concatenation in the t-product framework. As a result, the reconstructed global approximator \hat{f}_n in \mathbb{K}_c^d admits a unified network representation with uniform layer dimensions across all c slices, making it compatible with both the algebraic and geometric structure of the t-manifold.

Step 5): Uniform error control on the approximate t-manifold \mathcal{M}_ρ . To estimate the approximation error uniformly over the tubular neighborhood \mathcal{M}_ρ , fix any $x \in \mathcal{M}_\rho$. By definition of \mathcal{M}_ρ , there exists a closest point $\tilde{x} \in \mathcal{M}$ such that $\|x - \tilde{x}\|_{\mathbb{K}_c^d} \leq \rho$.

We decompose the total approximation error as follows:

$$\|f_0(x) - \hat{f}_n(x)\|_{\mathbb{K}_c} \leq \underbrace{\|f_0(x) - f_0(\tilde{x})\|_{\mathbb{K}_c}}_{(a)} + \underbrace{\|f_0(\tilde{x}) - \hat{f}_n(\tilde{x})\|_{\mathbb{K}_c}}_{(b)} + \underbrace{\|\hat{f}_n(\tilde{x}) - \hat{f}_n(x)\|_{\mathbb{K}_c}}_{(c)}. \quad (28)$$

Let us bound each term separately.

Term (a): Smoothness of f_0 . Since f_0 is C^β - \mathbb{K}_c -smooth on \mathcal{M} in the t-product sense according to (A2), its slice components $f_{0,k}$ satisfy

$$|[Mf_0(x)]_k - [Mf_0(\tilde{x})]_k| \leq B_0 \|[Mx]_k - [M\tilde{x}]_k\|_2 \leq B_0 \|x - \tilde{x}\|_{\mathbb{K}_c^d}.$$

Thus,

$$\|f_0(x) - f_0(\tilde{x})\|_{\mathbb{K}_c}^2 = \sum_{k=1}^c |[Mf_0(x)]_k - [Mf_0(\tilde{x})]_k|^2 \leq cB_0^2 \rho^2.$$

Term (b): Spectral approximation error. From the construction in Steps 3)-4), we have:

$$\|f_0(\tilde{x}) - \hat{f}_n(\tilde{x})\|_{\mathbb{K}_c}^2 = \sum_{k=1}^c |f_{0,k}([M\tilde{x}]_k) - f_{n,k}([M\tilde{x}]_k)|^2 \leq c\varepsilon^2.$$

Term (c): Lipschitz continuity of TNN approximation. The TNN approximator \hat{f}_n inherits Lipschitz smoothness from its ReLU subnetworks and the smooth A_k projections. For each k :

$$|f_{n,k}([Mx]_k) - f_{n,k}([M\tilde{x}]_k)| \leq L_k \cdot \|[Mx]_k - [M\tilde{x}]_k\|_2 \leq L_k \cdot \|x - \tilde{x}\|_{\mathbb{K}_c^d}.$$

Aggregating over k , we obtain: $\|\hat{f}_n(x) - \hat{f}_n(\tilde{x})\|_{\mathbb{K}_c}^2 \leq \sum_{k=1}^c L_k^2 \rho^2 \leq cL^2 \rho^2$.

Combining all three terms, we have:

$$\|f_0(x) - \hat{f}_n(x)\|_{\mathbb{K}_c}^2 \leq 3c (B_0^2 \rho^2 + \varepsilon^2 + L^2 \rho^2).$$

Now, we choose ρ such that $\rho^2 = \mathcal{O}(\varepsilon^2)$, i.e., set $\rho := C_0 \cdot \varepsilon$, where $\varepsilon = CB_0 d^{1/2} d_\delta^{3\beta/2} (W_0 L_0)^{-2\beta/d_\delta}$, so that all three terms become of the same order. This implies

$$\|f_0(x) - \hat{f}_n(x)\|_{\mathbb{K}_c}^2 \leq C' B_0^2 c (W_0 L_0)^{-4\beta/d_\delta},$$

uniformly for all $x \in \mathcal{M}_\rho$. \square

Step 3: Generalization Error for TNNs Having established the approximation proper of tensor neural networks on approximate t-manifolds in Lemma 12, we now quantify the stochastic error incurred by empirical risk minimization in the presence of Gaussian noise. We follow the covering number approach adapted from [29], and present a two-stage analysis.

Lemma 13 (Stochastic Error Bound for TNNs). *Let \mathcal{F}_n be a class of tensor neural networks with t -scalar parameters of total size \mathcal{S} and depth \mathcal{D} . Suppose each $f \in \mathcal{F}_n$ satisfies $\|f(x)\|_{\mathbb{K}_c} \leq \mathcal{B}$ for all $x \in \mathcal{M} \subset \mathbb{K}_c^d$, where $\mathcal{B} \geq B_0$ bounds the C^β -norm of f_0 . Assume the observations follow the Gaussian model $y_i = f_0(x_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 I_c)$. Then the empirical risk minimizer $\hat{f}_n \in \mathcal{F}_n$ satisfies:*

$$\mathbb{E}_S[L(f_0) - 2L_n(\hat{f}_n) + L(\hat{f}_n)] \leq C \cdot \mathcal{B}^5 \cdot \frac{SD \log(\mathcal{S})(\log n)^5}{n}, \quad (29)$$

provided $n \geq C_0 SD \log \mathcal{S}$.

Proof. The proof proceeds by adapting the frequency-slice truncation and symmetrization argument from [29] to the setting of \mathbb{K}_c^d -smooth functions under the t-product.

Step 1): Frequency Decomposition. Since $f_0(x), f(x), y$ are all in \mathbb{K}_c , we apply the orthogonal transform $M \in \mathbb{R}^{c \times c}$ defining the t-product:

$$\|y - f(x)\|_{\mathbb{K}_c}^2 = \sum_{k=1}^c |[M(y - f(x))]_k|^2 := \sum_{k=1}^c \ell_k(f, x, y).$$

We define $\mathcal{F}_{n,k} := \{[Mf(x)]_k : f \in \mathcal{F}_n\}$ as the function class for the k -th frequency slice.

Step 2): Truncation of Sub-Gaussian Responses. Define truncated responses:

$$T_{\beta_n}(z) := \begin{cases} z, & |z| \leq \beta_n, \\ \beta_n \cdot \text{sign}(z), & \text{otherwise,} \end{cases} \quad \text{with } \beta_n := C\mathcal{B} \log n. \quad (30)$$

For each slice k , define the truncated regression residual:

$$g_k^{\beta_n}(f, x, y) := (f_k([Mx]_k) - T_{\beta_n}([My]_k))^2 - (f_{0,k}([Mx]_k) - [T_{\beta_n}([My]_k)])^2.$$

We symmetrize the generalization gap:

$$G_k^{\beta_n}(f, x, y) := \mathbb{E}'[g_k^{\beta_n}(f, x', y')] - 2g_k^{\beta_n}(f, x, y).$$

Step 3): Bounding Expected Slice-Wise Gap. By [21, Theorem 11.4], we have for each k :

$$\mathbb{E}_S \left[\frac{1}{n} \sum_{i=1}^n G_k^{\beta_n}(f, x_i, y_i) \right] \leq C_1 \cdot \frac{\beta_n^4 \cdot \log \mathcal{N}_{2n}(1/n, \|\cdot\|_\infty, \mathcal{F}_{n,k}) \cdot (\log n)^4}{n}. \quad (31)$$

Step 4): Truncation Residual is Negligible. By [29, Lemma 3.2], the residual between the true and truncated loss satisfies:

$$|g_k(f, x, y) - g_k^{\beta_n}(f, x, y)| \leq C\beta_n |[My]_k| \cdot \mathbf{1}_{\{|[My]_k| > \beta_n\}} + C' \beta_n^2 e^{-c\beta_n^2}. \quad (32)$$

Since $[My]_k$ is sub-Gaussian, this term is exponentially small in β_n , and thus negligible.

Step 5): From Slice Networks to TNN Class. Each slice network $f_k \in \mathcal{F}_{n,k}$ is a scalar ReLU subnetwork of the full TNN $f \in \mathcal{F}_n$, sharing the same width \mathcal{W} and depth \mathcal{D} . Hence, from [2, Theorem 12.2]:

$$\log \mathcal{N}_{2n}(1/n, \|\cdot\|_\infty, \mathcal{F}_{n,k}) \leq C \cdot \text{Pdim}(\mathcal{F}_{n,k}) \cdot \log(\mathcal{B}n). \quad (33)$$

Moreover, by [4, Theorems 3 and 6], for ReLU networks:

$$\text{Pdim}(\mathcal{F}_{n,k}) \leq C' \cdot SD \log \mathcal{S}. \quad (34)$$

This is because the total parameter count \mathcal{S} and depth \mathcal{D} dominate the slice-wise complexity when all slices share the same architecture.

Step 6): Final Bound. Summing over $k = 1, \dots, c$ slices, we get:

$$\mathbb{E}_S[L(f_0) - 2L_n(\hat{f}_n) + L(\hat{f}_n)] \leq C'' \cdot c \cdot \frac{\mathcal{B}^5 SD \log \mathcal{S} \cdot (\log n)^5}{n}, \quad (35)$$

and absorbing c into C'' , the result follows. \square

We then combine this bound with the approximation error from Step 2 to obtain the final excess risk bound.

Lemma 14 (Expected Excess risk for TNNs in \mathbb{K}_c^d). *Let \mathcal{F}_n be the class of TNNs used to construct \hat{f}_n in Lemma 12, with t-scalar width \mathcal{W} , depth \mathcal{D} , and total parameter count \mathcal{S} determined by two positive integers W, L . Assume the noise model $y_i = f_0(x_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 I_c)$ and $x_i \in \mathcal{M}_\rho \subset \mathbb{K}_c^d$. Assume also that each $f \in \mathcal{F}_n$ satisfies $\|f(x)\|_{\mathbb{K}_c} \leq \mathcal{B}$ with $\mathcal{B} \geq B_0$. Then the empirical risk minimizer $\hat{f}_n \in \mathcal{F}_n$ satisfies:*

$$\mathbb{E}_S \left[\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \right] \leq C_1 \cdot \mathcal{B}^5 \cdot \frac{\mathcal{S} \cdot \mathcal{D} \log \mathcal{S} (\log n)^5}{n} + C_2 B_0^2 \cdot c \cdot (WL)^{-4\beta/d_\delta}, \quad (36)$$

provided $n \geq C_0 \cdot \mathcal{S} \mathcal{D} \log \mathcal{S}$.

This result follows by combining the approximation error bound from Lemma 12 with the stochastic error bound from Lemma 13, using the decomposition in Lemma 11.

Proof Sketch. By the approximation error bound in Lemma 12, for any $x \in \mathcal{M}_\rho$, we have

$$\|f_0(x) - \hat{f}_n(x)\|_{\mathbb{K}_c}^2 \leq C \cdot (WL)^{-4\beta/d_\delta}.$$

Next, applying the stochastic error bound in Lemma 13, we obtain:

$$\mathbb{E}_S [L(f_0) - 2L_n(\hat{f}_n) + L(\hat{f}_n)] \leq C \cdot \mathcal{B}^5 \cdot \frac{\mathcal{S} \cdot \mathcal{D} \log \mathcal{S} (\log n)^5}{n}.$$

Then, using risk decomposition (Lemma 11) completes the proof. \square

Step 4: Bias–Variance Tradeoff and Final Bound Having derived expected excess risk bound (Step 3), we now determine how to optimally allocate the network complexity, specifically the product of W and L controlling width and depth, to balance the two terms within Eq. (36).

We show that choosing WL in Lemma 14 appropriately as a function of sample size n leads to the optimal convergence rate. This step yields the final generalization bound for TNNs trained on data from an approximate t-manifold.

Lemma 15 (Bias–variance balance and effective t-manifold dimension). *The optimal tradeoff between approximation and stochastic bounds is achieved when $WL \asymp n^{d_\delta/(4\beta+2d_\delta)}$. Substituting into the generalization bound in Lemma 14 gives*

$$\mathbb{E}_S \left[\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \right] \leq C \cdot n^{-\frac{2\beta}{2\beta+p_{\text{eff}}}} \cdot (\log n)^5, \quad \text{where } p_{\text{eff}} := p \log(cd). \quad (37)$$

Proof. From Lemma 14, the error bound satisfies

$$\mathbb{E}_S \left[\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \right] \leq C_1 \cdot \mathcal{B}^5 \cdot \frac{\mathcal{S} \cdot \mathcal{D} \log \mathcal{S} \cdot (\log n)^5}{n} + C_2 \cdot B_0^2 \cdot c \cdot (WL)^{-4\beta/d_\delta}. \quad (38)$$

Recall from Lemma 12 that the network width and depth are given by $\mathcal{W} = \mathcal{O}(d_\delta^{\beta+1} W \log W)$, $\mathcal{D} = \mathcal{O}(L \log L)$, and total parameter count satisfies $\mathcal{S} = \mathcal{O}(\mathcal{W}^2 \cdot \mathcal{D}) = \mathcal{O}(d_\delta^{2\beta+2} W^2 \log^2 W \cdot L \log L)$.

Hence the stochastic error term behaves like

$$\frac{\mathcal{S} \cdot \mathcal{D} \log \mathcal{S}}{n} = \tilde{\mathcal{O}} \left(\frac{(WL)^2}{n} \right), \quad (39)$$

where $\tilde{\mathcal{O}}$ hides logarithmic factors and constants. Note that the approximation error term is $\mathcal{O}((WL)^{-4\beta/d_\delta})$.

To minimize the total error, we balance the two terms:

$$\frac{(WL)^2}{n} \asymp (WL)^{-4\beta/d_\delta} \quad \Rightarrow \quad WL \asymp n^{d_\delta/(4\beta+2d_\delta)}. \quad (40)$$

Substituting into Eq. (36) yields

$$\mathbb{E} \left[\|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \right] \leq \tilde{\mathcal{O}} \left(n^{-2\beta/(2\beta+d_\delta)} \right).$$

Finally, recall that the frequency-slice projection dimension satisfies $d_\delta = \mathcal{O}(p \log(cd))$ for fixed $\delta \in (0, 1)$, leading to the effective t-manifold dimension $p_{\text{eff}} := p \log(cd)$. This completes the proof. \square

E.2 Final Proof of the Main Theorem

We now complete the proof of the main generalization result by combining the four steps established above.

Proof of Theorem 4. The proof proceeds in four steps.

Step 1. We decompose the expected excess risk into a stochastic generalization term and an approximation error term (Lemma 11).

Step 2. We construct a TNN \hat{f}_n that uniformly approximates f_0 over the ρ -neighborhood \mathcal{M}_ρ with error at most $\mathcal{O}((WL)^{-4\beta/d_\delta})$ (Lemma 12).

Step 3. We control the stochastic error which is of order $\mathcal{O}(\mathcal{SD} \log \mathcal{S}(\log n)^5/n)$ (Lemma 13 and Lemma 14).

Step 4. By Lemma 15, we choose $WL \asymp n^{d_\delta/(4\beta+2d_\delta)}$ to balance the approximation and statistical errors. Substituting yields the optimal rate

$$\mathbb{E} \|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \leq \tilde{\mathcal{O}} \left(n^{-2\beta/(2\beta+p_{\text{eff}})} \right),$$

with $p_{\text{eff}} = \mathcal{O}(p \log(cd))$. This completes the proof. \square

We remark that Theorem 4 also recovers the exact manifold setting as a special case: when the input data lies precisely on the t-manifold \mathcal{M} (i.e., $\rho = 0$), the same generalization rate stated in Theorem 3 follows directly.

Discussion on Sharper Bounds on Exact t-Manifolds. If the signal is strictly supported on the t-manifold \mathcal{M} (i.e., $\rho = 0$), one can avoid the use of spectral projection and instead perform approximation entirely within the intrinsic manifold coordinates. Specifically, a sharper error bound is expected to be achieved through the following three-step strategy:

1. *Local Chart Construction:* For each frequency slice $\mathcal{M}_k = \{[Mx]_k : x \in \mathcal{M}\}$, construct a finite C^∞ atlas $\{(U_{\alpha,k}, \varphi_{\alpha,k})\}$ that covers \mathcal{M}_k , along with a subordinate smooth partition of unity $\{\eta_{\alpha,k}\}$.
2. *Slice-wise Approximation:* For each chart domain $\Omega_{\alpha,k} = \varphi_{\alpha,k}(U_{\alpha,k}) \subset \mathbb{R}^p$, approximate the localized function $f_{0,\alpha,k} := f_0 \circ \varphi_{\alpha,k}^{-1}$ using a ReLU network $\phi_{\alpha,k}$ with width $\mathcal{O}(\varepsilon^{-p/\beta})$ and depth $\mathcal{O}(\log(1/\varepsilon))$.
3. *TNN Assembly:* Combine the approximations via $\eta_{\alpha,k}$ to form $[M\hat{f}(x)]_k := \sum_\alpha \eta_{\alpha,k}([Mx]_k) \phi_{\alpha,k}(\varphi_{\alpha,k}([Mx]_k))$, and apply M^{-1} to reconstruct $\hat{f}(x)$.

This may yield a uniform approximation error $\|\hat{f}(x) - f_0(x)\|_{\mathbb{K}_c} \leq \varepsilon$ for all $x \in \mathcal{M}$, with total network complexity $\mathcal{O}(\varepsilon^{-p/\beta} \log(1/\varepsilon))$. Consequently, the resulting generalization rate is expected to be improved to

$$\mathbb{E} \|\hat{f}_n - f_0\|_{L^2(\nu)}^2 \leq \tilde{\mathcal{O}} \left(n^{-2\beta/(2\beta+p)} \right),$$

removing the logarithmic overhead from $p_{\text{eff}} = p \log(cd)$.

Nevertheless, achieving the sharpest possible convergence rate is not the focus of this paper. Our primary goal is to introduce a geometric learning framework for TNNs built upon the t-product algebra, and to develop a unified approximation and generalization theory for tensor-valued learning on non-Euclidean domains.

F Modeling Perspectives and Potential of t-Product Geometry

This section highlights the modeling potential of the t-product geometry introduced in this work. We discuss how its underlying algebraic and geometric structures give rise to constructs like t-modules and bi-module regularization, which can inform strategies for tensor-based data analysis. The BTR, presented in Section F.1, serves as a practical example, applying dual t-module constraints to tasks such as clustering and Poisson tensor completion. Experiments on varied data types, including images, video, hyperspectral imagery, point clouds, and thermal sequences, demonstrate BTR’s flexibility and effectiveness. Section F.2 further reflects on the broader relevance of t-product geometry in machine learning.

F.1 Example: Bidirectional Tensor Representation (BTR)

This subsection presents a detailed study of the BTR method, which illustrates how the proposed t-product geometry can inform practical modeling in visual computing tasks such as image clustering and video denoising. While the overall focus of this paper remains theoretical, BTR serves as an example of how abstract concepts from the t-manifold framework may be translated into data modeling principles.

From Theory to Practice: BTR as a Geometric Instantiation BTR draws directly from the structure of t-manifolds, which are defined through directional t-modules over the t-product algebra. In particular, BTR imposes simultaneous low-rankness in both the row-wise module \mathbb{K}_w^h and the column-wise module \mathbb{K}_h^w , reflecting the bi-module structure that underlies the non-flat geometry of general t-manifolds. This coupling introduces an interaction between orthogonal directions that departs from traditional flat models and provides a more expressive constraint for tensor recovery.

From this perspective, the BTR model can be viewed not only as a bidirectional extension of existing tensor low-dimensional representation approaches but also as a concrete example of how core elements of the proposed t-product geometry, such as t-linear structure, directional decomposition, and module-wise regularization, can be integrated into a unified modeling framework. While the scope of BTR remains limited to specific applications, it illustrates the potential of the geometric perspective introduced in this work to support the development of new learning models that are consistent with the underlying algebraic structure.

F.1.1 Motivation: From t-Module to Bidirectional Representation

The notion of a *t-module* provides a foundational building block in the theory of t-manifolds. It represents the most basic and flat instance of a manifold under the algebra of the t-product. Formally, a t-module is a free module over the ring \mathbb{K}_c induced by a fixed orthogonal transform (e.g., DFT or DCT), and comprises tensors expressible as t-linear combinations of a finite generating set. When a collection of images is encoded as a third-order tensor, their span under the t-product naturally forms a t-module. This structure offers a conceptually clean and algebraically consistent way to model self-expressiveness and low-dimensional structure in tensor data [83, 69, 68, 30].

Traditional tensor low-rank representation (t-LRR) methods [83] are fundamentally built upon this t-module perspective. Leveraging the t-product’s ability to encode transform-domain correlations, these methods decompose high-dimensional visual data, such as image sets or video frames, into low-rank components that preserve multi-way structure. Techniques based on the t-SVD have shown particular promise in applications like subspace clustering and tensor completion [83]. Their success lies in the use of the tensor nuclear norm (TNN)⁷, which effectively enforces low-rankness in the transform domain and aligns well with the spectral regularity of natural signals [83, 69, 68, 30].

Despite these strengths, existing t-LRR models exhibit a structural limitation in that they operate in a *unidirectional* manner. Specifically, correlations are modeled only along the row-wise mode, treating each image as a t-vector in \mathbb{K}_w^h . As a consequence, horizontal dependencies are emphasized, while vertical dependencies encoded along columns are largely ignored.

In visual data, this asymmetry can be consequential. Row-wise modeling captures horizontal alignment and continuity, but structural variations expressed primarily along the column direction,

⁷In this section, TNN refers to the *tensor nuclear norm*, rather than *tensor neural networks* as used elsewhere in the paper. We apologize for any confusion caused by this overloading of notation.

such as elongated or column-aligned patterns, remain underrepresented. As a result, representations learned in a single t-module tend to be anisotropic, which can degrade robustness when informative structure or distortions occur along the unmodeled dimension.

Key insights motivating BTR include:

- Traditional t-LRR models focus on a single t-module \mathbb{K}_w^h and neglect its symmetric counterpart \mathbb{K}_h^w . This design restricts modeling to row-wise dependencies and leads to anisotropic representations that underutilize column-wise structure.
- Although \mathbb{K}_w^h and \mathbb{K}_h^w are each flat (linear) modules in isolation, jointly constraining data to be low-dimensional in both induces **a coupling that geometrically “twists” the representation space**. This coupling can be interpreted as an effective, constraint-induced nonlinearity, which may offer advantages in approximation and generalization.

In response to these insights, the BTR formulation is proposed as a principled extension of the t-module viewpoint. BTR imposes simultaneous low-rank constraints in both the row-wise module \mathbb{K}_w^h and the column-wise module \mathbb{K}_h^w . By doing so, it allows both horizontal and vertical structures to contribute to the representation, thereby yielding a more balanced and expressive encoding of spatial dependencies.

Geometrically, the coupling of dual t-modules can be interpreted as introducing an effective form of twisting in the representation space: **although each module is flat in isolation, their joint low-dimensional constraint gives rise to nontrivial interactions that deviate from a single flat t-module**. This induced nonlinearity is subtle but appears beneficial in practice, enabling the model to capture complex visual patterns more faithfully.

We begin by applying BTR to the tensor low-rank representation problem [83], where it serves to model bidirectional dependencies in visual data. Building on this formulation, we then extend the same idea to the more challenging setting of Poisson tensor completion [78], demonstrating how bidirectional low-rank regularization can be adapted to photon-limited and sparsely observed data.

F.1.2 BTR Model for Tensor LRR

BTR formulates a generalized tensor low-rank representation model that simultaneously captures structural dependencies along both the row and column modes of visual data tensors. This formulation is grounded in the algebra of the t-product and systematically integrates bidirectional low-rank constraints with corruption modeling to enhance robustness and expressivity.

Let $\mathcal{X} \in \mathbb{R}^{h \times n \times w}$ denote a third-order data tensor, where each of the n frontal slices $\mathcal{X}(:, i, :)$ represents a grayscale image of size $h \times w$, with h and w denoting the height and width of each image, respectively. The objective is to decompose \mathcal{X} into low-rank components along both the row and column directions, while simultaneously estimating a sparse corruption tensor \mathcal{E} .

To this end, BTR solves the following optimization problem:

$$\begin{aligned} \min_{\mathcal{W}, \mathcal{R}, \mathcal{E}} \quad & \alpha \|\mathcal{W}\|_* + (1 - \alpha) \|\mathcal{R}\|_* + \lambda \Omega(\mathcal{E}), \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{D} * \mathcal{W} + \mathcal{E}, \quad \mathcal{X}^\perp = \mathcal{D}^\perp * \mathcal{R} + \mathcal{E}^\perp, \end{aligned} \tag{41}$$

where $\alpha \in [0, 1]$ and $\lambda > 0$ are regularization parameters. The other components in Problem (41) are interpreted as follows:

- $\mathcal{D} \in \mathbb{R}^{h \times p \times w}$ denotes the dictionary tensor in the row-wise direction. In the case of self-representation, which is commonly used in subspace clustering, we set $\mathcal{D} = \mathcal{X}$, following the design of prior works such as [68].
- $\mathcal{W} \in \mathbb{R}^{p \times n \times w}$ is the coefficient tensor capturing the linear relations between samples in the row-wise module \mathbb{K}_w^h . The representation $\mathcal{X} = \mathcal{D} * \mathcal{W} + \mathcal{E}$ encodes the t-linear combination of dictionary atoms to reconstruct \mathcal{X} along the row dimension, where $*$ denotes the t-product [68, 83].
- $\mathcal{X}^\perp = \text{permute}(\mathcal{X}, [3, 2, 1]) \in \mathbb{R}^{w \times n \times h}$ is the column-wise counterpart of the data tensor, obtained by cyclically permuting its modes to align the column dimension as the first mode. Similarly, $\mathcal{D}^\perp = \text{permute}(\mathcal{D}, [3, 2, 1])$ and $\mathcal{E}^\perp = \text{permute}(\mathcal{E}, [3, 2, 1])$ denote the permuted dictionary and error tensors for column-wise modeling.

- $\mathcal{R} \in \mathbb{R}^{p \times n \times h}$ is the coefficient tensor corresponding to the column-wise module \mathbb{K}_h^w , encoding the t-linear reconstruction of \mathcal{X}^\perp by \mathcal{D}^\perp . The constraint $\mathcal{X}^\perp = \mathcal{D}^\perp * \mathcal{R} + \mathcal{E}^\perp$ ensures that column-directional structures are also modeled with the same low-rank principle.
- $\|\cdot\|_\star$ denotes the tensor nuclear norm, which is the average of the nuclear norms of the frontal slices of the transformed tensor under M [41]. The use of tensor nuclear norm allows us to promote low-rank structure in the transform domain, which aligns with the inherent spectral regularity of image data.
- $\Omega(\mathcal{E})$ is a convex loss function for corruption modeling. Depending on the application scenario, we adopt either the element-wise ℓ_1 -norm $\|\mathcal{E}\|_1$ to promote sparsity in random corruptions [41], or the sample-wise ℓ_1 -norm $\|\mathcal{E}\|_{2,1}$ to promote robustness against structured outliers [82].

Compared to traditional t-LRR methods that rely solely on the row-wise formulation $\mathcal{X} = \mathcal{D} * \mathcal{W} + \mathcal{E}$ [83], the BTR model introduces a structurally consistent column-wise constraint via $\mathcal{X}^\perp = \mathcal{D}^\perp * \mathcal{R} + \mathcal{E}^\perp$. This dual-constraint structure ensures that both horizontal and vertical correlations in the data are simultaneously preserved, which is particularly important for visual data characterized by local spatial coherence in both dimensions.

By jointly minimizing the tensor nuclear norms of both \mathcal{W} and \mathcal{R} , the BTR formulation effectively learns a coupled low-rank structure that respects the bidirectional geometry of the underlying tensor space. This leads to improved robustness in tasks such as clustering and denoising, as the model better captures the intrinsic structure of the data while suppressing unstructured noise. From a geometric perspective, BTR can be viewed as simultaneously projecting the data onto two low-dimensional submodules in \mathbb{K}_w^h and \mathbb{K}_h^w , thereby enforcing global consistency and spectral regularity across both modes of the tensor.

Optimization Algorithm To efficiently solve the BTR optimization problem formulated in Equation (41), we adopt the Alternating Direction Method of Multipliers (ADMM) [50], which is particularly well-suited for problems involving multiple convex terms with linear constraints. ADMM decomposes the original constrained problem into simpler subproblems by introducing auxiliary variables, allowing for the separate optimization of nonsmooth objectives and linear operations.

To handle the coupled low-rank and reconstruction constraints, we introduce two auxiliary variables \mathcal{M} and \mathcal{N} corresponding to \mathcal{W} and \mathcal{R} , respectively. The reformulated optimization problem becomes (where the regularization $\Omega(\cdot)$ is instantiated by ℓ_1 -norm):

$$\begin{aligned} \min_{\mathcal{W}, \mathcal{R}, \mathcal{M}, \mathcal{N}, \mathcal{E}} \quad & \alpha \|\mathcal{M}\|_\star + (1 - \alpha) \|\mathcal{N}\|_\star + \lambda \|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{M} = \mathcal{W}, \quad \mathcal{N} = \mathcal{R}, \\ & \mathcal{X} = \mathcal{D} * \mathcal{W} + \mathcal{E}, \quad \mathcal{X}^\perp = \mathcal{D}^\perp * \mathcal{R} + \mathcal{E}^\perp. \end{aligned} \quad (42)$$

The augmented Lagrangian corresponding to the problem above is given by:

$$\begin{aligned} L = & \alpha \|\mathcal{M}\|_\star + (1 - \alpha) \|\mathcal{N}\|_\star + \lambda \|\mathcal{E}\|_1 \\ & + \langle \Lambda_1, \mathcal{M} - \mathcal{W} \rangle + \frac{\rho}{2} \|\mathcal{M} - \mathcal{W}\|_F^2 \\ & + \langle \Lambda_2, \mathcal{N} - \mathcal{R} \rangle + \frac{\rho}{2} \|\mathcal{N} - \mathcal{R}\|_F^2 \\ & + \langle \Lambda_3, \mathcal{X} - \mathcal{D} * \mathcal{W} - \mathcal{E} \rangle + \frac{\rho}{2} \|\mathcal{X} - \mathcal{D} * \mathcal{W} - \mathcal{E}\|_F^2 \\ & + \langle \Lambda_4, \mathcal{X}^\perp - \mathcal{D}^\perp * \mathcal{R} - \mathcal{E}^\perp \rangle + \frac{\rho}{2} \|\mathcal{X}^\perp - \mathcal{D}^\perp * \mathcal{R} - \mathcal{E}^\perp\|_F^2, \end{aligned} \quad (43)$$

where $\Lambda_1, \Lambda_2, \Lambda_3$, and Λ_4 are the Lagrange multiplier tensors associated with each constraint, and $\rho > 0$ is the penalty parameter controlling the convergence rate and constraint enforcement. The notation $\langle \cdot, \cdot \rangle$ denotes the standard inner product over third-order tensors, and all Frobenius norms are evaluated over the entire tensor volume.

Block-wise Updates. The ADMM algorithm proceeds by alternately minimizing L with respect to each primal variable block, followed by gradient ascent on the dual variables. Each subproblem admits a closed-form solution or a well-defined proximal update:

(1) *Update of \mathcal{M} and \mathcal{N}* :

$$\mathcal{M}^{(k+1)} = \text{Prox}_{\frac{\alpha}{\rho}}^{\|\cdot\|_*} \left(\mathcal{W}^{(k)} - \frac{1}{\rho} \Lambda_1^{(k)} \right), \quad \mathcal{N}^{(k+1)} = \text{Prox}_{\frac{1-\alpha}{\rho}}^{\|\cdot\|_*} \left(\mathcal{R}^{(k)} - \frac{1}{\rho} \Lambda_2^{(k)} \right). \quad (44)$$

These updates apply the tensor singular value thresholding (t-SVT) operator to promote low-rankness in the transform domain [42]. The proximal mapping of the tensor nuclear norm is computed by applying singular value soft-thresholding slice-wise after transforming via M (e.g., DFT).

(2) *Update of \mathcal{E}* : To simultaneously satisfy the row-wise and column-wise reconstruction constraints, we define:

$$\mathcal{A} = \mathcal{X} - \mathcal{D} * \mathcal{W}^{(k)} + \frac{1}{\rho} \Lambda_3^{(k)}, \quad \mathcal{B} = \text{permute} \left(\mathcal{X}^\perp - \mathcal{D}^\perp * \mathcal{R}^{(k)} + \frac{1}{\rho} \Lambda_4^{(k)}, [3, 2, 1] \right), \quad (45)$$

and compute the update as:

$$\mathcal{E}^{(k+1)} = \text{Prox}_{\frac{\lambda}{2\rho}}^{\|\cdot\|_1} \left(\frac{\mathcal{A} + \mathcal{B}}{2} \right). \quad (46)$$

This update applies the proximal operator of the ℓ_1 norm [41], encouraging element-wise sparsity.

(3) *Update of \mathcal{W} and \mathcal{R}* : These steps correspond to least-squares updates involving t-product linear systems:

$$\mathcal{W}^{(k+1)} = (\mathcal{I} + \mathcal{D}^\top * \mathcal{D})^{-1} * \left(\mathcal{M}^{(k+1)} + \frac{1}{\rho} \Lambda_1^{(k)} + \mathcal{D}^\top * (\mathcal{X} - \mathcal{E}^{(k+1)}) + \frac{1}{\rho} \mathcal{D}^\top * \Lambda_3^{(k)} \right), \quad (47)$$

$$\mathcal{R}^{(k+1)} = (\mathcal{I} + \mathcal{D}^{\perp\top} * \mathcal{D}^\perp)^{-1} * \left(\mathcal{N}^{(k+1)} + \frac{1}{\rho} \Lambda_2^{(k)} + \mathcal{D}^{\perp\top} * (\mathcal{X}^\perp - \mathcal{E}^\perp) + \frac{1}{\rho} \mathcal{D}^{\perp\top} * \Lambda_4^{(k)} \right). \quad (48)$$

Efficient implementations rely on the block diagonalization of t-product matrices in the transform domain, reducing the computational complexity.

(4) *Dual variable updates*:

$$\Lambda_1^{(k+1)} = \Lambda_1^{(k)} + \rho \left(\mathcal{M}^{(k+1)} - \mathcal{W}^{(k+1)} \right), \quad (49)$$

$$\Lambda_2^{(k+1)} = \Lambda_2^{(k)} + \rho \left(\mathcal{N}^{(k+1)} - \mathcal{R}^{(k+1)} \right), \quad (50)$$

$$\Lambda_3^{(k+1)} = \Lambda_3^{(k)} + \rho \left(\mathcal{X} - \mathcal{D} * \mathcal{W}^{(k+1)} - \mathcal{E}^{(k+1)} \right), \quad (51)$$

$$\Lambda_4^{(k+1)} = \Lambda_4^{(k)} + \rho \left(\mathcal{X}^\perp - \mathcal{D}^\perp * \mathcal{R}^{(k+1)} - \mathcal{E}^{(k+1)\perp} \right). \quad (52)$$

Convergence and Complexity. This two-block ADMM formulation enjoys provable convergence properties for convex objectives. The dominant cost per iteration lies in the t-SVD operations used in proximal mappings and t-product inverses, with per-iteration complexity approximately $\mathcal{O}(h n w \log w + h n^2 w)$ assuming DFT is used for the t-product. In practice, convergence is declared when the relative change in variables or objective falls below a tolerance (e.g., 10^{-6}), typically achieved in tens of iterations.

Summary of Procedure. The full algorithm is presented below.

Algorithm 1 ADMM for BTR Optimization

- 1: **Input:** Data tensor \mathcal{X} , dictionary \mathcal{D} , parameters α, λ, ρ
 - 2: **Initialize:** $\mathcal{W}, \mathcal{R}, \mathcal{M}, \mathcal{N}, \mathcal{E}$, and multipliers $\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4$
 - 3: **repeat**
 - 4: Update $\mathcal{M} \leftarrow \text{Prox}_{\frac{\alpha}{\rho}}^{\|\cdot\|_*}(\mathcal{W} - \rho^{-1} \Lambda_1)$
 - 5: Update $\mathcal{N} \leftarrow \text{Prox}_{\frac{1-\alpha}{\rho}}^{\|\cdot\|_*}(\mathcal{R} - \rho^{-1} \Lambda_2)$
 - 6: Compute \mathcal{A}, \mathcal{B} from current \mathcal{W}, \mathcal{R}
 - 7: Update $\mathcal{E} \leftarrow \text{Prox}_{\frac{\lambda}{2\rho}}^{\|\cdot\|_1}((\mathcal{A} + \mathcal{B})/2)$
 - 8: Update \mathcal{W}, \mathcal{R} via closed-form t-product linear solves
 - 9: Update $\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4$ via dual ascent
 - 10: **until** convergence criteria satisfied
 - 11: **Output:** Optimized variables $\mathcal{W}, \mathcal{R}, \mathcal{E}$
-

Experimental Settings for BTR-TLRR To evaluate the empirical effectiveness and generality of BTR, we conduct experiments on two representative visual computing tasks: image clustering and tensor recovery (video denoising). These tasks are selected to reflect the complementary strengths of BTR in unsupervised structure discovery and robust low-rank modeling under corruption.

Image Clustering. This task aims to cluster a set of unlabeled face or digit images into their respective categories based on structural similarities. The BTR model is applied in a self-representation setting, where each image is reconstructed from others using both row-wise and column-wise t-linear combinations.

Datasets. We use five benchmark datasets with varying numbers of classes and resolutions:

- *FRDUE*: 3040 facial images from 152 individuals, each of size 25×22 .
- *FRDUE-100*: A subset of FRDUE with 2500 images from 100 subjects, same resolution.
- *Olivetti*: 575 face images from 20 subjects, each of size 56×46 .
- *PIE-10*: 680 images of 10 people under varying lighting/pose, size 22×22 .
- *USPS1000*: 1000 handwritten digits from USPS dataset, 10 classes, 16×16 resolution.

Preprocessing. All images are first normalized to the range $[0, 1]$ to ensure numerical stability and consistent scaling. Each image is then reshaped into a tensor slice $\mathcal{X}(:, i, :) \in \mathbb{R}^{h \times 1 \times w}$ and stacked along the sample axis to form the data tensor $\mathcal{X} \in \mathbb{R}^{h \times n \times w}$. No additional augmentation or feature extraction is performed, as we aim to evaluate the capacity of BTR to directly exploit raw spatial correlations.

Parameter Settings. The dictionary tensor is set as $\mathcal{D} = \mathcal{X}$ to enable self-representation. The bidirectional regularization weight is fixed as $\alpha = 0.5$ to balance row-wise and column-wise contributions. The regularization parameter λ is selected from $\{0.1, 0.5, 1, 2, 5\}$ based on validation performance. The penalty parameter ρ in ADMM is initialized to 1 and updated adaptively during iterations. The stopping criterion is set to a relative objective change less than 10^{-6} .

Evaluation Metrics. We use three standard clustering quality measures:

1. *Accuracy (ACC)*: The maximum matching accuracy between predicted and true labels.
2. *Normalized Mutual Information (NMI)*: Measures the mutual information normalized by entropy.
3. *Purity (PUR)*: Proportion of correctly assigned samples per cluster.

Clustering is performed using spectral clustering with normalized cuts (Ncut) on the affinity matrix \hat{Z} obtained by symmetrizing the sum of the row-wise and column-wise coefficient tensors. Each clustering result is averaged over 10 independent runs to account for randomness in initialization. The results are shown in Table 1.

Tensor Recovery (Video Denoising). In this task, the objective is to recover clean video sequences from corrupted observations by exploiting low-rank structure in the tensorized video frames. BTR is employed to jointly denoise in both spatial directions.

Datasets. We use seven standard YUV video sequences with diverse content, such as *Coastguard*, *Foreman*, *Mother and Daughter*, etc. Each sequence is cropped and converted to grayscale with a typical resolution of 144×176 and 100–300 frames. The resulting tensor has dimensions $\mathcal{X} \in \mathbb{R}^{w \times n \times h}$, where the second mode indexes frames.

Preprocessing. Videos are first converted to grayscale and scaled to pixel range $[0, 255]$. Frames are reshaped as $\mathcal{X}(:, i, :) \in \mathbb{R}^{w \times 1 \times h}$ and stacked to form the full data tensor. Corruption is simulated by replacing 30% of randomly selected pixels with uniformly random noise in $[0, 255]$.

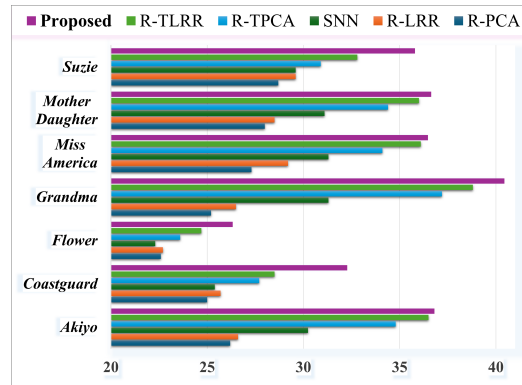


Figure F.2: PSNR of video denoising on YUV sequences.

Parameter Settings. The dictionary is again set as $\mathcal{D} = \mathcal{X}$ for self-representation. We fix $\alpha = 0.5$ and choose $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1\}$ by grid search. The penalty parameter is initialized as $\rho = 0.1$ and dynamically adjusted. The convergence threshold is set to 10^{-5} . Only the ℓ_1 -norm is used here, as pixel-level corruption is assumed to be sparse and unstructured.

Evaluation Metric. Reconstruction quality is assessed by Peak Signal-to-Noise Ratio (PSNR), computed as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{255^2}{\text{MSE}} \right),$$

where MSE is the mean squared error between the recovered and clean frames. Higher PSNR indicates better recovery. The results⁸ are shown in Figure F.2.

F.1.3 BTR Model for Poisson Tensor Completion

In this subsection, we extend our proposed BTR to the challenging setting of *Poisson tensor completion*, which commonly arises in photon-limited imaging applications [78].

Poisson Tensor Completion and Motivation Let $\mathcal{X}^* \in \mathbb{R}_+^{h \times n \times w}$ denote the underlying clean tensor to be recovered, with h and w corresponding to spatial dimensions and n indexing frames or spectral bands. Due to photon limitations or sensor constraints, only a subset of entries is observed, and the measurements are corrupted by signal-dependent Poisson noise [78, 8].

Inspired by the geometric formulation developed in this work, we propose to regularize the recovery process using the t-product structure along both spatial directions. Specifically, we model \mathcal{X}^* as lying near the intersection of two low-dimensional t-modules, \mathbb{K}_w^h and \mathbb{K}_h^w , capturing row- and column-wise regularities. This bi-module perspective, rooted in the t-manifold geometry introduced earlier, leads naturally to a new regularization—the Bi-Module-based Tensor Nuclear Norm (BTR-TNN)—for Poisson tensor completion.

Observation Model Formally, the observed tensor $\mathcal{Y} \in \mathbb{N}_+^{h \times n \times w}$ is modeled as [78]:

$$\mathcal{Y}_{ijl} = \mathcal{B}_{ijl} \cdot \mathfrak{P}(\mathcal{X}_{ijl}^* + c), \quad \forall (i, j, l) \in [h] \times [n] \times [w], \quad (53)$$

where:

- $\mathcal{B}_{ijl} \in \{0, 1\}$ is a binary sampling mask indicating whether entry (i, j, l) is observed;
- $\mathfrak{P}(\lambda)$ denotes a Poisson random variable with intensity parameter λ ;
- $c > 0$ is a known background constant accounting for ambient signal or electronic bias.

That is, when $\mathcal{B}_{ijl} = 1$, the observation \mathcal{Y}_{ijl} is drawn from a Poisson distribution with mean $\mathcal{X}_{ijl}^* + c$:

$$\mathbb{P}[\mathcal{Y}_{ijl} = m] = \frac{(\mathcal{X}_{ijl}^* + c)^m}{m!} \exp(-(\mathcal{X}_{ijl}^* + c)), \quad \forall m \in \mathbb{N}_+. \quad (54)$$

In the context of tensor geometry, the observation model (53) induces a sampling operator on the underlying t-module signal, where entrywise masking is applied to \mathcal{X}^* followed by pointwise Poisson corruption. This modeling is essential in practical systems where physical sensors (e.g., photon detectors) produce discrete, count-based measurements with signal-dependent variance.

Throughout this work, we also assume the sampling tensor \mathcal{B} follows a uniform Bernoulli distribution [78]:

Assumption 2 (Uniform Bernoulli Sampling). Each entry is independently observed with probability $p \in (0, 1]$. That is,

$$\mathcal{B}_{ijl} \sim \text{Bernoulli}(p), \quad \text{i.i.d. over } i \in [h], j \in [n], l \in [w].$$

This assumption facilitates theoretical analysis and algorithmic derivation. However, as we later show, our formulation and optimization algorithm remain valid under more general (non-uniform and structured) sampling schemes, which are tested in the experiments.

⁸As the compared TLRR used DFT as the linear transform, the proposed BTR also uses DFT in video recovery.

Bi-Module-based Poisson Tensor Completion Model To recover the underlying tensor \mathcal{X}^* from the incomplete and noisy observations \mathcal{Y} , we adopt a maximum likelihood estimation framework. Due to the Poisson statistics of the measurement process, the negative log-likelihood function of any candidate tensor $\mathcal{X} \in \mathbb{R}_+^{h \times n \times w}$ takes the form:

$$\mathcal{L}(\mathcal{X}; \mathcal{Y}, \mathcal{B}) := \sum_{i=1}^h \sum_{j=1}^n \sum_{l=1}^w \mathcal{B}_{ijl} [(\mathcal{X}_{ijl} + c) - \mathcal{Y}_{ijl} \log(\mathcal{X}_{ijl} + c)], \quad (55)$$

where the constant $c > 0$ is added to ensure bounded likelihood and avoid degeneracy at $\mathcal{X}_{ijl} = 0$.

Since the tensor \mathcal{X}^* is assumed to admit simultaneous low-rank structures in both the row-wise and column-wise t-modules, we impose a spectral regularization based on the Bi-Module-based Tensor Nuclear Norm (BTR-TNN). Specifically, for a tensor $\mathcal{X} \in \mathbb{R}^{h \times n \times w}$, define:

$$\mathcal{X}_{\langle 1 \rangle} := \mathcal{X}, \quad \mathcal{X}_{\langle 2 \rangle} := \text{permute}(\mathcal{X}, [3, 2, 1]),$$

where $\mathcal{X}_{\langle 1 \rangle}$ aligns with the row-wise t-module \mathbb{K}_w^h and $\mathcal{X}_{\langle 2 \rangle}$ with the column-wise t-module \mathbb{K}_h^w . We then define:

$$\|\mathcal{X}\|_{\text{BTR}} := \alpha_1 \|\mathcal{X}_{\langle 1 \rangle}\|_{\star} + \alpha_2 \|\mathcal{X}_{\langle 2 \rangle}\|_{\star}, \quad \text{with } \alpha_1 + \alpha_2 = 1, \alpha_i \geq 0. \quad (56)$$

This BTR-TNN regularization promotes low-rankness simultaneously along both tensor modules, offering a symmetric modeling of structural dependencies. Combining the data-fidelity term and structural prior, we formulate the regularized BTR model for Poisson tensor completion as:

$$\hat{\mathcal{X}} \in \arg \min_{\mathcal{X} \in \mathcal{C}} \lambda \cdot \mathcal{L}(\mathcal{X}; \mathcal{Y}, \mathcal{B}) + \|\mathcal{X}\|_{\text{BTR}}, \quad (57)$$

where $\lambda > 0$ is the regularization parameter and \mathcal{C} is a spikiness-constrained feasible set:

$$\mathcal{C} := \{\mathcal{X} \in \mathbb{R}_+^{h \times n \times w} \mid \mathcal{X}_{ijl} \leq \beta \text{ for all } i, j, l\}.$$

The set \mathcal{C} captures the natural assumption that real-world signals are bounded and non-negative. The optimization problem (57) thus seeks a tensor that both fits the Poisson observations and admits bidirectional spectral compressibility in the transform domain. In the next section, we propose an efficient ADMM-based algorithm to solve this problem.

Optimization Algorithm The key idea is to decouple the BTR-TNN regularization from the Poisson log-likelihood via auxiliary variables.

Let us introduce two auxiliary tensors $\mathcal{K}^{(1)}$ and $\mathcal{K}^{(2)}$, corresponding to the two permuted modules $\mathcal{X}_{\langle 1 \rangle}$ and $\mathcal{X}_{\langle 2 \rangle}$ respectively. We then rewrite Eq. (57) as the following constrained problem:

$$\begin{aligned} \min_{\mathcal{X}, \{\mathcal{K}^{(i)}\}} \quad & \lambda \cdot \mathcal{L}(\mathcal{X}; \mathcal{Y}, \mathcal{B}) + \sum_{i=1}^2 \alpha_i \|\mathcal{K}_{\langle i \rangle}^{(i)}\|_{\star} \\ \text{s.t.} \quad & \mathcal{K}^{(i)} = \mathcal{X}, \quad \forall i \in \{1, 2\}, \quad \mathcal{X} \in \mathcal{C}, \end{aligned} \quad (58)$$

where each $\mathcal{K}_{\langle i \rangle}^{(i)}$ denotes a permuted version of $\mathcal{K}^{(i)}$ aligned with the i -th t-module.

We define the forward and inverse permutations as:

$$\mathfrak{F}_1(\mathcal{X}) := \mathcal{X}, \quad \mathfrak{F}_2(\mathcal{X}) := \text{permute}(\mathcal{X}, [3, 2, 1]), \quad \mathfrak{F}_i^{-1} := \mathfrak{F}_i.$$

We also introduce an auxiliary variable \mathcal{Z} to enforce the spikiness constraint explicitly. The augmented Lagrangian becomes:

$$\begin{aligned} \mathcal{L}_{\rho}(\mathcal{X}, \{\mathcal{K}^{(i)}\}, \mathcal{Z}, \mathcal{V}, \{\mathcal{W}^{(i)}\}) := & \lambda \cdot \mathcal{L}(\mathcal{X}; \mathcal{Y}, \mathcal{B}) + \sum_{i=1}^2 \alpha_i \|\mathfrak{F}_i(\mathcal{K}^{(i)})\|_{\star} + \mathbb{I}_{\mathcal{C}}(\mathcal{X}) \\ & + \frac{\rho}{2} \|\mathcal{Z} - \mathcal{X} + \rho^{-1} \mathcal{V}\|_F^2 + \sum_{i=1}^2 \frac{\rho}{2} \|\mathcal{K}^{(i)} - \mathcal{Z} + \rho^{-1} \mathcal{W}^{(i)}\|_F^2, \end{aligned}$$

where $\mathcal{V}, \{\mathcal{W}^{(i)}\}$ are the dual variables, $\rho > 0$ is a penalty parameter, and $\mathbb{I}_{\mathcal{C}}$ is the indicator function of the constraint set \mathcal{C} .

The ADMM updates alternate over three blocks: (1) \mathcal{Z} , (2) \mathcal{X} and $\{\mathcal{K}^{(i)}\}$, and (3) dual variables. Each step admits closed-form or efficiently computable updates using proximal operators of the nuclear norm and KL-divergence terms. The detailed subproblem solutions and convergence criteria are presented in Algorithm 2.

Algorithm 2 ADMM for BTR-based Poisson Tensor Completion

- 1: **Input:** Observation tensor \mathcal{Y} , sampling mask \mathcal{B} , background c , regularization parameter λ , BTR weights α_1, α_2 , penalty parameter ρ , bound β , tolerance ϵ
- 2: **Initialize:** $\mathcal{X}_0 = \mathcal{Z}_0 = \mathbf{0}$, $\mathcal{K}_0^{(1)} = \mathcal{K}_0^{(2)} = \mathbf{0}$, $\mathcal{W}_0^{(1)} = \mathcal{W}_0^{(2)} = \mathbf{0}$, $\mathcal{V}_0 = \mathbf{0}$
- 3: **for** $t = 0, 1, 2, \dots$ until convergence **do**
- 4: **Update** primal variable \mathcal{Z} :

$$\mathcal{Z}_{t+1} = \Pi_{\mathcal{C}} \left(\frac{1}{3} \left(\mathcal{X}_t + \rho^{-1} \mathcal{V}_t + \sum_{i=1}^2 \mathcal{K}_t^{(i)} + \rho^{-1} \mathcal{W}_t^{(i)} \right) \right)$$

- 5: **Update** primal variable \mathcal{X} by solving:

$$\mathcal{X}_{t+1} = \arg \min_{\mathcal{X} \in \mathbb{R}_+^{h \times n \times w}} \lambda \cdot \mathcal{L}(\mathcal{X}; \mathcal{Y}, \mathcal{B}) + \frac{\rho}{2} \|\mathcal{X} - (\mathcal{Z}_{t+1} + \rho^{-1} \mathcal{V}_t)\|_F^2$$

- 6: **Update** each $\mathcal{K}^{(i)}$ via singular value thresholding:

$$\mathcal{K}_{t+1}^{(i)} = \mathfrak{F}_i^{-1} \left(\text{Prox}_{\alpha_i/\rho}^{\|\cdot\|_*} \left(\mathfrak{F}_i(\mathcal{Z}_{t+1} + \rho^{-1} \mathcal{W}_t^{(i)}) \right) \right)$$

- 7: **Update dual variables:**

$$\mathcal{V}_{t+1} = \mathcal{V}_t + \rho(\mathcal{Z}_{t+1} - \mathcal{X}_{t+1}), \quad \mathcal{W}_{t+1}^{(i)} = \mathcal{W}_t^{(i)} + \rho(\mathcal{Z}_{t+1} - \mathcal{K}_{t+1}^{(i)})$$

- 8: **Check convergence:**

$$\max \left\{ \|\mathcal{X}_{t+1} - \mathcal{X}_t\|_{\infty}, \|\mathcal{Z}_{t+1} - \mathcal{Z}_t\|_{\infty}, \max_i \|\mathcal{K}_{t+1}^{(i)} - \mathcal{K}_t^{(i)}\|_{\infty} \right\} < \epsilon$$

- 9: **and**

$$\|\mathcal{Z}_{t+1} - \mathcal{X}_{t+1}\|_{\infty} < \epsilon, \quad \max_i \|\mathcal{Z}_{t+1} - \mathcal{K}_{t+1}^{(i)}\|_{\infty} < \epsilon$$

10: **end for**

11: **Output:** Estimated tensor $\hat{\mathcal{X}} = \mathcal{X}_{t+1}$

Experimental Results To evaluate the effectiveness of the proposed BTR model under Poisson noise, we conduct simulated Poisson tensor completion on a variety of datasets from visual computing and remote sensing. Our goal is to validate that bidirectional low-rank modeling via BTR-TNN leads to more robust and accurate recovery under photon-limited sampling.

We consider two sampling protocols for constructing the binary mask tensor \mathcal{B} :

1. *Uniform sampling with ratio p* : Each entry of the underlying tensor is observed independently with probability $p \in (0, 1]$ as described in Assumption 2.
2. *Non-uniform sampling with proxy ratio p* : A more challenging setting where entire rows, columns, or tubes may be missing. Specifically, we sequentially apply structured dropout along each mode with probability $q = (1 - p)^{1/3}$, such that the expected observed fraction is approximately p .

For each ground-truth tensor $\mathcal{X}^* \in \mathbb{R}_+^{h \times n \times w}$, we normalize \mathcal{X}^* such that $\|\mathcal{X}^*\|_{\infty} = \beta = 100$ and set the background constant $c = 5$. A Poisson corrupted tensor \mathcal{P} is generated entrywise from

$$\mathcal{P}_{ijl} \sim \mathfrak{P}(\mathcal{X}_{ijl}^* + c),$$

followed by masking to obtain $\mathcal{Y} = \mathcal{B} \circledast \mathcal{P}$, where \circledast denotes the entrywise (Hadamard) product.

We compare the proposed BTR model against a diverse set of state-of-the-art baselines for Poisson tensor completion:

1. *PoissonNN* [8]: Matrix nuclear norm minimization under Poisson noise on unfolded tensors.
2. *HaLRTC* [38]: Tucker-based tensor completion via high-accuracy low-rank minimization.
3. *TMac* [72]: Parallel matrix factorization over unfolded modes.

4. *TNN-DFT/DCT* [78]: Tensor nuclear norm based completion with different linear transforms in the t-SVD (DFT and DCT).

We implement two variants of our proposed method *BTR-DFT* and *BTR-DCT* using DFT and DCT, respectively.

We evaluate our method on four categories of real-world tensor data commonly encountered in remote sensing and visual computing:

- *Hyperspectral Images (HSI)*: We use two widely studied hyperspectral datasets from AVIRIS: *Indian Pines* ($145 \times 145 \times 224$) and *Salinas A* ($86 \times 83 \times 224$). For computational efficiency, we retain the first 30 spectral bands and reshape them into tensors of size $h \times n \times w$, where $h = 145$ or 86 , $n = 30$, and $w = 145$ or 83 respectively. These tensors represent spatial-spectral cubes widely used in target detection and classification tasks.
- *Multispectral Images (MSI)*: We include *Cloth* and *Oil Painting* from the Columbia MSI Database⁹, each resized to $256 \times 256 \times 31$. These data capture rich material textures under multiple spectral bands, supporting fine-grained texture recovery.
- *LiDAR Point Cloud (Depth and Intensity)*: We use 30 consecutive frames from the *Scenario B* sequence in the Velodyne HDL-64E dataset¹⁰. The data are formatted as two tensors of size $64 \times 30 \times 250$, one for distance and the other for intensity. These tensors model 3D point clouds with missing strips due to occlusion or view constraints.
- *Infrared Datasets*. We evaluate our method on two thermal infrared datasets that reflect different sensing scenarios. The first is Sequence 1 from the OSU Thermal Database¹¹ [12], consisting of 30 frames of human activity under low illumination at a resolution of 320×240 , stacked into a tensor of size $320 \times 30 \times 240$. The second is from the Infrared Detection dataset [25], which includes dim-small aircraft targets under complex backgrounds. We use the first 30 frames from data3.zip¹², forming a tensor of size $256 \times 30 \times 256$.

All methods are quantitatively evaluated using PSNR and SSIM metrics. For each sampling setting, experiments are repeated over 10 independent trials, and the average results are reported. The complete quantitative comparisons are summarized in Tables F.3–F.6. For qualitative analysis, representative visual results are provided in Figures F.3 and F.4.

Table F.3: Comparison of Poisson tensor completion performance on the *IndiaPines* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	7.77	0.009	8.54	0.017	22.65	0.399	23.8	0.442
HaLRTC	23.07	0.408	22.91	0.416	23.13	0.437	22.96	0.441
TMac	24.21	0.548	24.65	0.582	24.25	0.553	24.65	0.584
TNN-DFT	24.65	0.407	24.82	0.422	25.4	0.454	25.52	0.471
TNN-DCT	25.9	0.504	26.59	0.531	27.21	0.583	27.82	0.606
BTR-DFT	28.73	0.72	29.38	0.727	29.37	0.746	29.97	0.749
BTR-DCT	28.44	0.715	29.65	0.762	29.03	0.739	30.34	0.787

Summary of Experimental Results Across all evaluated datasets and sampling scenarios, our proposed BTR demonstrates consistently improved performance over existing tensor completion baselines. In particular, BTR-DCT and BTR-Data variants achieve the highest PSNR and SSIM scores in most cases, especially under challenging conditions such as non-uniform sampling.

These results highlight the practical benefits of incorporating bidirectional low-rank priors via the Bi-Module-based Tensor Nuclear Norm. By jointly regularizing both the row-wise and column-wise

⁹<http://www1.cs.columbia.edu/CAVE/databases/multispectral>

¹⁰<http://www.mrt.kit.edu/z/publ/download/velodynettracking/dataset.html>

¹¹<https://www.cse.ohio-state.edu/otcbvs-bench/Bench/IRdata/>

¹²<http://www.csdata.org/p/387/>

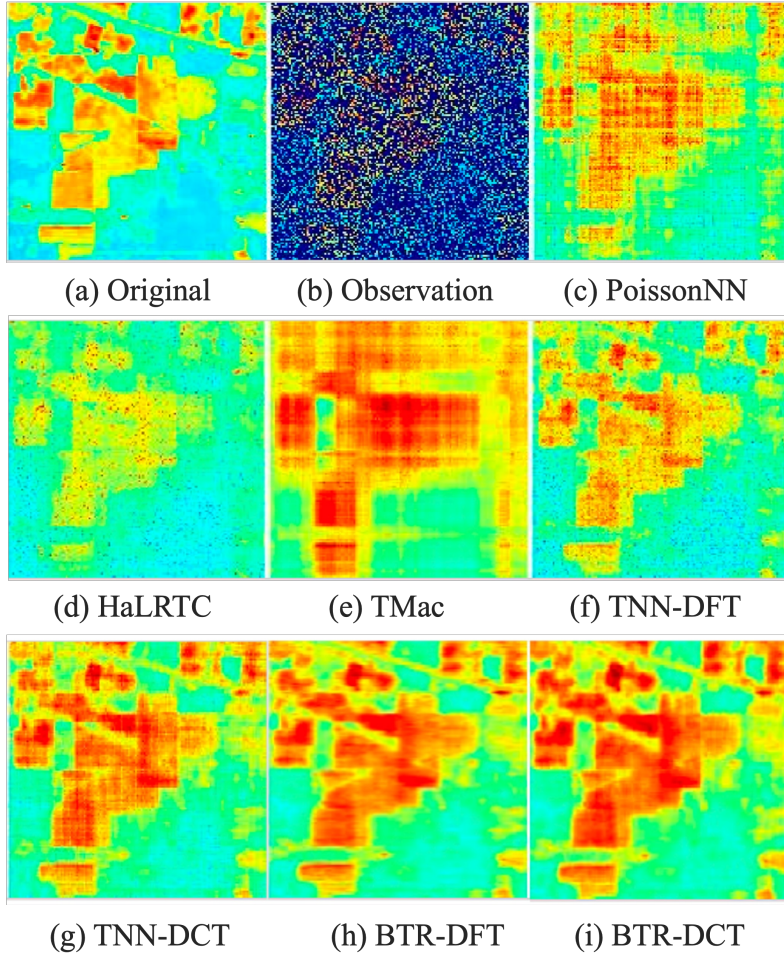


Figure F.3: A visual example of Poisson tensor completion on a selected spectral band of the *Indian Pines* dataset under the *uniform sampling setting*, where 20% of entries are observed independently at random. The observed elements follow Poisson distributions parameterized by the ground-truth intensity values. Twelve methods are compared, including PoissonNN [8], HaLRTC [38], TMac [72], TNN-DFT [78], TNN-DCT [78], and two proposed variants of the BTR model.

Table F.4: Comparison of Poisson tensor completion performance on the *SalinasA* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	7.62	0.034	8.4	0.049	21.41	0.445	23.39	0.5
HaLRTC	23.13	0.451	23.17	0.441	23.77	0.484	23.46	0.469
TMac	25.01	0.672	25.39	0.689	25.01	0.67	25.37	0.685
TNN-DFT	21.74	0.364	22.28	0.386	24.93	0.475	25.56	0.49
TNN-DCT	25.75	0.552	26.92	0.596	27.88	0.636	29.11	0.663
BTR-DFT	28.72	0.759	30.48	0.808	28.91	0.774	30.75	0.816
BTR-DCT	29.73	0.767	31.48	0.816	29.73	0.765	31.58	0.815

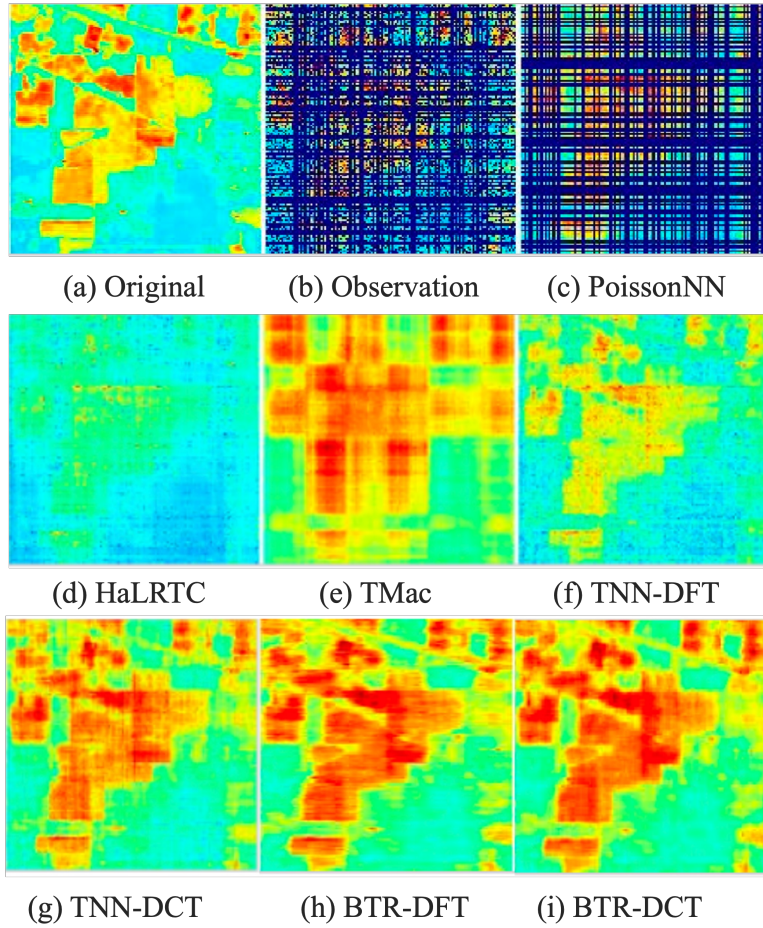


Figure F.4: A visual example of Poisson tensor completion on a selected spectral band of the *Indian Pines* dataset under the *non-uniform sampling setting*, where 20% of entries are observed. The observed elements follow Poisson distributions parameterized by the ground-truth intensity values. Twelve methods are compared, including PoissonNN [8], HaLRTC [38], TMac [72], TNN-DFT [78], TNN-DCT [78], and two proposed variants of the BTR model.

Table F.5: Comparison of Poisson tensor completion performance on the *Cloth* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	13.36	0.09	13.36	0.145	20.84	0.38	21.35	0.472
HaLRTC	21.2	0.445	21.53	0.529	21.42	0.473	21.93	0.566
TMac	21.11	0.417	21.83	0.492	21.11	0.417	21.85	0.492
TNN-DFT	22.93	0.569	23.36	0.654	25.84	0.694	26.2	0.761
TNN-DCT	24.57	0.63	25.24	0.726	26.91	0.742	27.73	0.82
BTR-DFT	25.71	0.721	26.63	0.755	28.55	0.823	29.32	0.832
BTR-DCT	25.96	0.752	27.16	0.791	28.94	0.853	30.15	0.868

Table F.6: Comparison of Poisson tensor completion performance on the *OilPainting* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	15.11	0.068	13.95	0.136	23.64	0.326	21.98	0.45
HaLRTC	23.33	0.339	22.3	0.458	23.7	0.377	22.91	0.495
TMac	22.58	0.353	21.38	0.452	22.59	0.35	21.38	0.455
TNN-DFT	27.52	0.421	26.33	0.572	28.69	0.464	27.68	0.612
TNN-DCT	28.4	0.466	27.23	0.624	29.59	0.511	28.63	0.667
BTR-DFT	28.96	0.554	29.57	0.552	30.68	0.596	31.23	0.588
BTR-DCT	29.72	0.638	30.66	0.633	31.22	0.671	32.25	0.662

Table F.7: Comparison of Poisson tensor completion performance on the *SenerioBDis* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	13.24	0.077	14.08	0.123	16.28	0.163	17.48	0.23
HaLRTC	17.61	0.218	18.75	0.263	19.28	0.27	20.3	0.308
TMac	18.54	0.228	18.89	0.253	18.53	0.227	18.88	0.252
TNN-DFT	20.11	0.355	21.4	0.385	22.73	0.417	23.66	0.438
TNN-DCT	20.05	0.391	21.41	0.421	22.69	0.46	23.73	0.478
BTR-DFT	20.18	0.412	21.12	0.431	21.62	0.465	22.86	0.482
BTR-DCT	20.13	0.449	21.09	0.467	22.04	0.521	23.24	0.529

Table F.8: Comparison of Poisson tensor completion performance on the *SenerioBIntens* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	14.19	0.12	14.89	0.196	16.75	0.239	17.95	0.331
HaLRTC	18.06	0.327	19.01	0.4	18.81	0.375	19.92	0.448
TMac	17.76	0.269	18.1	0.297	17.77	0.27	18.12	0.298
TNN-DFT	20.45	0.469	21.36	0.528	22.3	0.559	23.21	0.597
TNN-DCT	20.47	0.477	21.41	0.537	22.25	0.57	23.27	0.61
BTR-DFT	21.06	0.531	21.93	0.566	22.23	0.584	23.22	0.619
BTR-DCT	20.65	0.511	21.58	0.559	22.16	0.599	23.18	0.634

Table F.9: Comparison of Poisson tensor completion performance on the *Infraed* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	7.59	0.002	8.34	0.002	27.46	0.177	27.65	0.175
HaLRTC	24.94	0.138	24.1	0.128	24.6	0.146	23.91	0.137
TMac	31.66	0.42	31.94	0.428	31.67	0.418	31.93	0.427
TNN-DFT	27.28	0.111	26.57	0.11	27.31	0.112	26.62	0.116
TNN-DCT	29.93	0.188	29.28	0.17	30.0	0.188	29.35	0.175
BTR-DFT	30.69	0.246	29.97	0.226	30.92	0.263	30.11	0.231
BTR-DCT	32.33	0.427	32.73	0.392	32.61	0.442	32.88	0.397

Table F.10: Comparison of Poisson tensor completion performance on the *Thermal* dataset under various sampling configurations. Bold values highlight the best PSNR or SSIM in each setting. Gray cells indicate models proposed in this paper.

Model	NonUniform, 20%		NonUniform, 30%		Uniform, 20%		Uniform, 30%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PoissonNN	10.53	0.072	11.29	0.114	17.1	0.255	19.18	0.351
HaLRTC	19.29	0.381	20.64	0.439	20.57	0.438	21.86	0.492
TMac	15.46	0.21	15.83	0.245	15.46	0.209	15.83	0.245
TNN-DFT	25.66	0.579	26.59	0.592	27.94	0.629	28.18	0.624
TNN-DCT	26.01	0.62	27.42	0.639	28.69	0.679	29.5	0.678
BTR-DFT	26.08	0.657	27.23	0.664	29.09	0.705	29.62	0.695
BTR-DCT	26.5	0.708	27.88	0.718	29.11	0.748	30.31	0.746

Table F.11: Potential applications of t-product geometry in machine learning. Each entry connects an application area with corresponding tools and expected modeling advantages.

Application Domain	t-Geometric Tools	Potential Modeling Advantages
Representation Learning	t-modules	Respects multi-mode structure, avoids flattening, supports spectral-aware representations
Low-Rank Regularization	Dual modules $\mathbb{K}_w^h, \mathbb{K}_h^w$	Captures richer spatial-spectral dependencies; enhances robustness to anisotropic noise
Manifold Optimization	t-gradients, t-Grassmann structures	Enables first-order optimization over transform-constrained low-rank tensor manifolds
Graph-Based Learning	t-Laplacian, frequency slices	Supports localized, frequency-adaptive graph structures for structured propagation
Generative Modeling	Low-dimensional t-manifolds	Imposes spectral priors for generating structured tensor data near coherent subspaces
Spectral Manifold Learning	Transform-sliced t-geometry	Allows nonlinear embedding and kernel design directly in frequency-transformed domains

t-modules, BTR captures richer structural correlations in the transform domain, which are often under-exploited by unidirectional models such as standard TNN or matrix-based methods.

The consistent advantage of BTR suggests that bidirectional modeling serves as a more robust and geometrically faithful inductive bias for tensor recovery in Poisson noise settings, particularly when spatial and spectral dependencies interact nontrivially.

F.2 Potential Applications of t-Product Geometry in Machine Learning

This work establishes a geometric framework for tensor analysis based on the algebra of the t-product. Key constructs such as t-modules, bi-module regularity, and t-manifolds formalize the notion of directional low-rank structure in third-order tensors. While the focus is theoretical, these constructs open up several potential modeling paradigms across different areas of machine learning, especially those involving multi-dimensional data and spectral structure.

Structured Representation Learning. In data modalities such as hyperspectral imaging, EEG, and multiview video, the observations inherently form third-order tensors, where different modes encode complementary information. Conventional approaches often flatten such data, which can destroy

spatial, temporal, or spectral coherence. Modeling with t-modules preserves this multi-mode structure and allows representation learning to proceed directly in the transform domain. In this domain, local dependencies and alignments can be more naturally captured, potentially leading to semantically meaningful representations in unsupervised settings or neural architectures that incorporate tensorized latent spaces [48, 52].

Bidirectional Regularization. Traditional low-rank regularization is typically applied along one tensor mode and may fail to capture richer inter-mode interactions. The introduction of dual-mode regularization through bi-modules, namely, imposing low-rankness in both \mathbb{K}_w^h and \mathbb{K}_h^w , allows the model to adapt to both spatial and spectral variations. This form of regularization supports more robust structural recovery, especially when dealing with anisotropic data where noise or sparsity patterns are direction-dependent. Applications such as Poisson completion, denoising, and sparse recovery may benefit from this enriched formulation [78, 54].

Optimization Over t-Structured Manifolds. The t-product framework induces structured low-rank tensor spaces that resemble matrix manifolds, such as Grassmannian or Stiefel manifolds, but defined in the transform domain. This structure supports the development of optimization algorithms that operate over these sets using generalized notions of gradients, retractions, and local coordinate charts defined slice-wise. Such optimization strategies may be advantageous in tasks involving tensor completion, constrained learning, or structured regression where low-rank constraints are essential [19, 59].

Graph Modeling in Tensor Domains. Graph-based methods play a central role in modeling local neighborhood structure and propagating information. When applied to tensor data, however, standard graph constructions may fail to capture directional or spectral dependencies. The t-Laplacian operator enables the construction of frequency- and slice-aware graphs, where connectivity is modulated by spectral behavior. This provides a pathway toward localized, adaptive graph learning that reflects structural heterogeneity and temporal variation, with potential applications in spatiotemporal forecasting, multimodal sensor fusion, and spectral graph neural networks [70, 44, 51].

Geometry-Aware Generative Modeling Many generative frameworks, such as VAEs or diffusion models, assume latent spaces with simple isotropic geometry. However, real-world data often exhibit structured variability confined to low-dimensional subspaces. The concept of low-dimensional t-manifolds introduces an alternative by allowing the generation process to respect spectral coherence and mode-wise constraints. By generating samples near structured subspaces in the transform domain, one may improve coherence and interpretability of generated content. While this idea remains exploratory, it aligns with recent trends in incorporating geometric structure into generative modeling [11, 60, 14].

Manifold Learning in the Transform Domain Classical manifold learning methods define local neighborhoods and similarity graphs based on Euclidean distances in the ambient space. This perspective may be suboptimal for tensor data where local geometry varies across frequency or structural modes. The t-product framework allows defining neighborhoods and affinities directly in the transform slices, enabling spectral variants of manifold learning. In this setting, geodesics and local charts can be computed per slice, potentially reducing embedding distortion and enhancing stability to sampling noise. This approach may benefit nonlinear embedding, clustering, and visualization tasks, especially when robustness and interpretability are desired [84].

Although these modeling directions remain preliminary, they are grounded in the algebraic and geometric formulation introduced in this work. The t-product geometry provides a unified perspective for analyzing, optimizing, and generating structured tensor data. By explicitly accounting for directional structure, spectral coherence, and transform-domain constraints, this framework may serve as a foundation for future research in tensor-based representation learning and model design.

As a natural next step, the development of open-source software libraries and computational toolkits that implement t-geometric operations such as spectral neighborhood construction, transform-slice optimization, manifold-regularized learning, and generative model adaptation will be important for enabling broader experimentation, reproducibility, and practical deployment in real-world applications.