Consistent Geometric Deep Learning via Hilbert Bundles and Cellular Sheaves

Kartik Tandon[1,*,†] **Julian Gould**[2,*] **Tanishq Bhatia**[3]
Francesca Dominici[4] **Alejandro Ribeiro**[1] **Claudio Battiloro**[5,4,*,†]

[1]University of Pennsylvania [2]Sakana AI [3]Northeastern University [4]Harvard University [5]Brown University

[*]Equal contribution

[†]Corresponding authors: ktandon@sas.upenn.edu, claudio_battiloro@brown.edu

Abstract

Modern deep learning architectures increasingly contend with sophisticated signals that are natively infinite-dimensional, such as time series, probability distributions, or operators, and are defined over irregular domains. Yet, a unified learning theory for these settings has been lacking. To start addressing this gap, we introduce a novel convolutional learning framework for possibly infinite-dimensional signals supported on a manifold. Namely, we use the connection Laplacian associated with a Hilbert bundle as a convolutional operator, and we derive filters and neural networks, dubbed as *HilbNets*. We make HilbNets and, more generally, the convolution operation, implementable via a two-stage sampling procedure. First, we show that sampling the manifold induces a Hilbert Cellular Sheaf, a generalized graph structure with Hilbert feature spaces and edge-wise coupling rules, and we prove that its sheaf Laplacian converges in probability to the underlying connection Laplacian as the sampling density increases. Notably, this result is a generalization to the infinite-dimensional bundle setting of the Belkin & Niyogi [14] convergence result for the graph Laplacian to the manifold Laplacian, a theoretical cornerstone of geometric learning methods. Second, we discretize the signals and prove that the discretized (implementable) HilbNets converge to the underlying continuous architectures and are transferable across different samplings of the same bundle, providing consistency for learning. Finally, we validate our framework on synthetic and real-world tasks. Overall, our results broaden the scope of geometric learning as a whole by lifting classical Laplacian-based frameworks to settings where the signal at each point lives in its own Hilbert space.

1 Introduction

Over the past few years, advances in deep learning have delivered state-of-the-art performance across many areas, driven by increasingly expressive architectures and corresponding gains in both theory and practice. A major contributor to this success, though not the only one, has been the rise of Convolutional Neural Networks (CNNs) [63]. CNNs have shown outstanding results in settings ranging from image recognition [60] to speech processing [1]. At their core, CNNs rely on filters leveraging the regular (often metric) organization of common signal types, such as spatial grids. In contrast, many modern datasets live on irregular, non-Euclidean domains, including social networks for detection and recommendation [2] or point clouds for shape segmentation [107], to name only a few. Such structured data can be represented by richer mathematical objects, among which networks and manifolds are prominent. Motivated by this, the intuition behind CNNs has been generalized to graph convolutional neural networks (GCNs) [86, 41, 59] and extended to many other settings, e.g. simplicial complexes [10, 17, 6], cell complexes [43, 16, 77], order lattices [84], and manifolds [101, 34, 87, 11, 27]. Nevertheless, existing works do not address convolutional filtering of infinite-dimensional signals on manifolds, despite such data being ubiquitous in practice, from time series and spatiotemporal fields arising in sensing, robotics, and climate science to distributional and measure-valued representations common in modern learning systems [54].

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

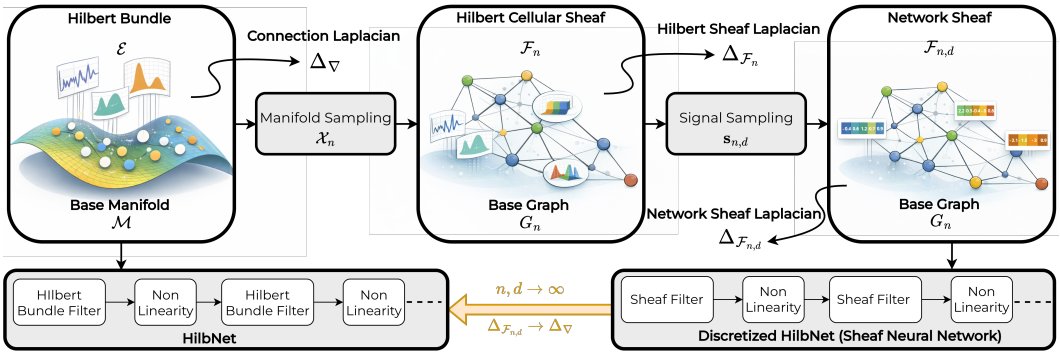

Figure 1: **Overview of the HilbNets framework.** A HilbNet is a convolutional neural network processing infinite-dimensional signals supported on $\mathcal{M}$ (e.g., time-series or distributions over curved domains). The convolutional operator is the connection Laplacian Δ_∇. To make HilbNets implementable, we first sample n points $\mathcal{X}_n$ from $\mathcal{M}$ to obtain a Hilbert Cellular Sheaf $\mathcal{F}_n$ with associated Hilbert Sheaf Laplacian $\Delta_{\mathcal{F}_n}$. We then take d samples of the signals to get vectors $\mathbf{s}_{n,d} \in \mathbb{R}^{nd}$ living on a network sheaf $\mathcal{F}_{\mathcal{X}_{n,d}}$, i.e., a generalized matrix-weighted graph, with associated network Sheaf Laplacian $\Delta_{\mathcal{F}_{n,d}} \in \mathbb{R}^{nd \times nd}$. Discretized HilbNets are then Sheaf Neural Networks that take as inputs $\mathbf{S}_{n,d}$ and $\Delta_{\mathcal{F}_{n,d}}$. We prove that Discretized HilbNets converge to the underlying HilbNet as the number of manifold and signal samples goes to infinity.

To address this gap, we adopt a bundle viewpoint. Informally, a bundle $\mathcal{E}$ over a base manifold $\mathcal{M}$ is a consistent assignment to each point $x \in \mathcal{M}$ of a space $\mathcal{E}_x$, called a fiber. A *section* is a map that picks an element $S(x) \in \mathcal{E}_x$ at every point. In other words, signals supported on manifolds can be seen as sections, e.g., scalar manifold signals correspond to $\mathcal{E}_x \simeq \mathbb{R}$ [101], or tangent bundle signals correspond to $\mathcal{E}_x \simeq T_x\mathcal{M}$ [11]. In this work, we develop a convolutional learning framework operating over *Hilbert bundles*, i.e., bundles whose fibers are (possibly) infinite-dimensional Hilbert spaces. We design a bundle-theoretic convolutional learning framework and, to make it implementable, we draw the first rigorous connection between Hilbert bundles and *Hilbert Cellular Sheaves*, generalized graph structures whose nodes and edges carry infinite-dimensional signals along with consistency rules.

Related Works. The connection between continuous domains, such as manifolds and bundles, and discrete structures, such as graphs and cellular sheaves, first emerged in pioneering investigations on the so-called manifold hypothesis. This hypothesis posits that, although data may live in a high-dimensional ambient space, they are effectively generated by sampling from one or several low-dimensional Riemannian manifolds [38]. The manifold hypothesis underpins several modern spectral graph methods, e.g., nonlinear dimensionality-reduction, clustering, interpretability, and learning algorithms that exploit latent geometric structures. The renowned work of Belkin and Niyogi [14] proved that, assuming access to a finite point cloud sampled from the underlying manifold, it is possible to build a weighted undirected graph whose Laplacian converges to the Laplace-Beltrami operator of the underlying manifold in probability as the number of samples goes to infinity.

The work in [14], and related results, e.g., [92, 93], have been used, directly or indirectly, to design learning systems over manifolds and networks [102, 67, 20, 74, 11, 56]. Despite the diversity of such systems, these models all assume finite-dimensional fibers and therefore do not directly address learning with infinite-dimensional manifold signals. The main technical reason behind this gap is the lack of an extension of the convergence result in [14] to bundles with infinite-dimensional fibers.

A line of related works of interest comes from cellular sheaf theory. Cellular sheaves are combinatorial instances of sheaves introduced in [90] and later rediscovered in [29]. In [18, 50, 7, 39, 37, 78], neural networks operating on finite-dimensional cellular sheaves over graphs, referred to as *network sheaves*, are presented, generalizing graph neural networks by, intuitively, replacing scalar edge weights with learned or structured matrix weights. Recently, the works in [9, 11] showed that neural networks for tangent bundle signals can be implemented as certain sheaf neural networks operating on network sheaves built from manifold samples. For an extended treatment of related work, see Appendix A.

Contribution. Our contributions are as follows:

C1. We first define a ***convolution operation over a Hilbert bundle*** through its associated *connection Laplacian*. This convolution extends Laplacian-based convolutions on tangent bundles [11], manifolds [102], and graphs [91, 40], as well as standard time convolutions. Using the Borel functional calculus,

we then define ***Hilbert bundle convolutional filters*** for infinite-dimensional manifold signals. These filters are general and expressive, and can be instantiated through suitable spectral responses.

C2 We then introduce ***HilbNets***, deep convolutional architectures whose layers stack Hilbert bundle filters and pointwise nonlinearities. HilbNets are continuous models and are therefore not directly implementable. To address this, we provide a principled discretization of the manifold domain by sampling points and showing that the induced structure is a *Hilbert cellular sheaf* over an undirected graph. The corresponding sheaf Laplacian combines scalar edge weights, obtained from the sampled base manifold, with parallel transport maps associated with the bundle geometry or learned from data. We prove that this sheaf Laplacian converges in probability to the connection Laplacian as the sampling density increases, yielding the first extension of the classical convergence result of [14] to the infinite-dimensional bundle setting. We then discretize the signals themselves to obtain an implementable architecture, show that discretized HilbNets are novel instances of network sheaf neural networks, and prove that they converge to the corresponding continuous architectures as both the manifold and signal sampling densities increase. Moreover, we show that discretized HilbNets are *transferable* across different samplings of the same underlying bundle, providing resolution consistency guarantees for learning.

C3 Finally, we show the effectiveness of HilbNets on a synthetic transport recovery task and on real-world traffic forecasting tasks, comparing them against baselines with different inductive biases in order to isolate the benefits of the bundle formulation.

The potential impact of this work extends well beyond the HilbNet architecture. See Appendix B for a detailed discussion of broader impact and future directions, and Fig. 1 for an overview.

2 Preliminaries

Signals on Manifolds. Given a manifold $\mathcal{M}$, a vector-valued signal is a square-integrable function $S \in L^2(\mathcal{M}, \mathbb{R}^n)$. Certain vector-valued signals on $\mathcal{M}$ may possess the richer structure of a vector field, i.e., they are *sections* of the tangent bundle $T\mathcal{M}$ of $\mathcal{M}$ and thus elements of $L^2(\mathcal{M}, T\mathcal{M})$. More generally, we may consider signals that are L^2-sections of an arbitrary *bundle* $\mathcal{E}$. A bundle is called trivial when, for a generic fiber $\mathcal{V}$, it can be written as a product $\mathcal{E} = \mathcal{M} \times \mathcal{V}$. In this setting, $L^2(\mathcal{M}, \mathbb{R}^n)$ may be understood as the space of sections of the *trivial bundle* $\mathcal{E} = \mathcal{M} \times \mathbb{R}^n$. Consider now the case where the signal is 'infinite-dimensional', for instance, representing a time series recorded at each point $x \in \mathcal{M}$. While this is usually considered as a function $S : \mathcal{M} \times \mathbb{R} \to \mathbb{R}^n$, it may instead be more richly understood as a section of a *Hilbert bundle*, i.e., a bundle whose fibers are Hilbert spaces. As we will see, Hilbert bundles provide a principled and versatile approach to incorporating structural properties of infinite-dimensional data.

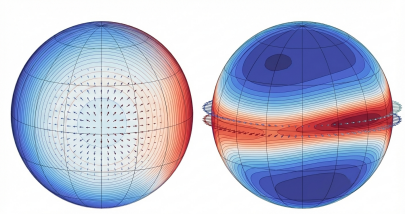

Figure 2: Visualization of the effect of the choice of underlying connection for generating heat flows of vector fields on the sphere $\mathbb{S}^2$. Left: Generated by considering the standard Levi-Civita connection on $T\mathbb{S}^2$, and corresponding Laplacian Δ_∇. Right: Generated by allowing a more general connection ∇ that allows for torsion anisotropy and considering the corresponding connection Laplacian Δ_∇.

Example 1. *In physics, Hilbert bundles often arise naturally when considering global geometric properties of quantum mechanical systems [4].*

Example 2. *In information geometry, the key objects of study are manifolds $\mathcal{M}$ given by the underlying parameters of some family of data distributions. This manifold is then equipped with a Riemannian structure by either the Otto-Wasserstein or Fisher-Rao metric, the latter of which locally recovers KL divergence. The proper analogue of the tangent bundle in this setting is a Hilbert bundle [72].*

Convolution, Heat Equation, and Connection Laplacian. Geometric signal processing and deep learning [66, 22] traditionally aim to develop convolutional filters and neural networks designed to respect the underlying geometry of the signals of interest. The relevant convolutional operators can usually be realized as a *connection Laplacian* $\Delta_\nabla : L^2(\mathcal{M}, \mathcal{E}) \to L^2(\mathcal{M}, \mathcal{E})$ operator realized from a connection ∇. For instance, for the tangent bundle over the circle $T\mathbb{S}^1$, the eigenfunctions of Δ_∇, with ∇ the Levi-Civita connection, recover the usual Fourier basis. Similarly, the eigenfunctions of Δ_∇ for the tangent bundle of the sphere $T\mathbb{S}^2$ recover spherical harmonics. Thus, convolutions with the connection Laplacian may be understood as generalized Fourier transforms in the spectral domain. In the spatial domain, it can be seen as performing a geometry-aware 'local averaging' of a signal

over fibers. Formally, the connection Laplacian is the generator of the heat equation in $L^2(\mathcal{M}, \mathcal{E})$,

$$\frac{\partial \mathbf{U}(x,t)}{\partial t} = -\Delta_\nabla \mathbf{U}(x,t), \tag{1}$$

where $\mathbf{U}(x,t)$ is the distribution of heat at $\mathcal{E}_x$ for $x \in \mathcal{M}$ at time $t \in \mathbb{R}_+$. A subtlety of note is that for non-Euclidean spaces, there is typically no canonical identification between fibers $\mathcal{E}_x$ and $\mathcal{E}_y$. Intuitively, the connection ∇ precisely encodes a globally coherent notion of transport between fibers. That is, inducing *parallel transport* maps $P_\gamma : \mathcal{E}_{\gamma(0)} \to \mathcal{E}_{\gamma(1)}$ for a path $\gamma \subset \mathcal{M}$, allowing us to compare elements across fibers along this path. More formally, the connection is used to define a first-order ODE whose solution is given by parallel transport (see Appendix G.1.3). The connection Laplacian is then the self-adjoint operator $\Delta_\nabla := \nabla^*\nabla$, now more clearly understandable as a 'local weighted average' over fibers with 'weights' corresponding to our choice of parallel transport. A more rigorous introduction to the relevant mathematical background is provided in Appendix G.

Example 3. *As seen in Fig. 2, the choice of connection can be used to emphasize aspects of the geometry that may be relevant for a particular task. For instance, most PDE-based approaches to color-image regularization can be realized as heat equations for a suitable choice of connection [8].*

3 Hilbert Bundle Filters and Neural Networks

In this section, we develop a convolutional learning framework for infinite-dimensional data, such as time series or probability distributions, indexed by a manifold $\mathcal{M}$. Core objects are *Hilbert bundles*.

Hilbert Bundles. Given a closed Riemannian manifold $\mathcal{M}$, a Hilbert bundle $\mathcal{E}$ over $\mathcal{M}$ is a bundle whose potentially infinite-dimensional fibers are separable Hilbert spaces over $\mathbb{R}$. The assumption of real, instead of complex, Hilbert spaces is not essential to our analysis, and is made only for the sake of exposition. As mentioned in Section 2, a *Hilbert Bundle signal* is then an L^2-section $S : \mathcal{M} \to \mathcal{E}$. Integration of sections in this setting should be understood in the Bochner integral sense, a generalized notion of integration for functions whose values lie in a Hilbert space rather than in $\mathbb{R}^n$. In finite dimensions, it reduces to the standard Lebesgue integral. Given fibers $\mathcal{H}_x$ and $\mathcal{H}_y$ of the Hilbert bundle $\mathcal{E}$, we consider *unitary* parallel transport operators $P_{x\to y} : \mathcal{H}_x \to \mathcal{H}_y$. As before, a globally compatible collection of such transport operators determines a connection ∇, with the subtlety that derivatives of sections must now be understood in the Fréchet sense. Intuitively, Fréchet differentiability is the infinite-dimensional analogue of ordinary differentiability: it asks that a section admit a best linear approximation under small perturbations, but where the linear approximation acts between Hilbert spaces. We therefore refer to this construction as a *Fréchet connection*, which recovers the usual notion of connection and covariant derivative when restricted to finite-dimensional bundles. As before, we obtain a self-adjoint operator $\Delta_\nabla := \nabla^*\nabla$ on $L^2(\mathcal{M}, \mathcal{E})$. Unlike the finite-dimensional case, however, this operator need not be compact and thus need not possess a discrete spectrum. As such, care must be taken when adapting classical arguments that involve spectral properties of the Laplacian to the Hilbert-bundle setting. Formal definitions of Hilbert bundles and Fréchet connections are provided in Appendix G. For a triple $(\mathcal{M}, \mathcal{E}, \nabla)$, where ∇ is a choice of Fréchet connection on the Hilbert bundle $\mathcal{E}$ over the manifold $\mathcal{M}$, we now wish to construct a general notion of a 'filtering' operation using the connection Laplacian Δ_∇. In finite-dimensional or compact settings, filters are often defined by applying a function directly to the eigenvalues of the Laplacian. In our setting, the appropriate analogue of eigenvalue-by-eigenvalue filtering is furnished by the *Borel functional calculus*, which allows us to apply a filter to a self-adjoint operator by instead integrating over its spectral measure. See Appendix G.4 for details.

Definition 1 (Hilbert bundle convolutional filter)**.** *A **convolutional filter** is specified by a bounded compactly supported Borel function $g \in L^\infty_c(\mathbb{R})$. The filtering of a signal $S \in L^2(\mathcal{M}, \mathcal{E})$ is then its convolution $\star_{\Delta_\nabla}$ with g defined as $g \star_{\Delta_\nabla} S := g(\Delta_\nabla)S$, where $g(\Delta_\nabla) : L^2(\mathcal{M}, \mathcal{E}) \to L^2(\mathcal{M}, \mathcal{E})$ is the bounded linear operator obtained by applying g to Δ_∇ through the Borel functional calculus.*

In this sense, g is the learnable frequency response, as in spectral graph neural filters [40], except we now use the spectral measure of the connection Laplacian acting on Hilbert bundle-valued signals.

Definition 2 (Hilbert bundle convolutional neural network)**.** *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle. A Hilbert bundle convolutional neural network, or **HilbNet**, is specified by a filter bank $\mathcal{W} = \{g^\ell_{u,q}\}_{\ell,u,q}$ with $g^\ell_{u,q} \in L^\infty_c(\mathbb{R})$, and a Lipschitz continuous nonlinear activation $\sigma : \mathbb{R} \to \mathbb{R}$. Given input signals $S_1, ..., S_{F_0} \in L^2(\mathcal{M}, \mathcal{E})$, the L-layer network output is obtained by the recursion*

$$S^{\ell+1}_u = \sigma\left(\sum_{q=1}^{F_\ell} g^\ell_{u,q}(\Delta_\nabla) S^\ell_q\right), \qquad \ell = 0, \dots, L-1, \qquad S^0_q = S_q, \tag{2}$$

where σ is applied pointwise in each fiber.

We concisely denote a HilbNet with $\Omega(\mathcal{E}, \Delta_\nabla, \mathcal{W}, \sigma)$. Similarly to the finite-dimensional case, a nonlinear activation $\sigma : \mathbb{R} \to \mathbb{R}$ with $\sigma(0) = 0$ extends to an operator on the L^2 section by simply picking a basis and then applying σ to each coordinate with respect to the chosen basis. It is straightforward to check that, for each $\ell \in \{0, \dots, L\}$, the layer signal S^ℓ remains an L^2 section.

4 Discretized HilbNets via Cellular Sheaves

HilbNets are continuous architectures that cannot be implemented directly in practice. Moreover, we typically do not have access to the true bundle and connection structure, but only to a point cloud or graph sampled from the underlying manifold $\mathcal{M}$, together with samples of the signal at each point or node. In this section, we first analyze the *Hilbert cellular sheaf* induced by spatial, i.e., manifold-level, sampling. We then further discretize the fibers, i.e., the signal domain itself, obtaining a finite rank *network sheaf*. This two-stage sampling is the basis of our consistency theory presented in Section 5. It allows us to prove that the fully discrete (thus, implementable) Laplacian and HilbNet converge to their infinite-dimensional counterparts, and hence that learning is consistent across scales.

Manifold Sampling. A generalized viewpoint to the theory of bundles is given by the language of *sheaves*, mathematical structures initially introduced by Jean Leray while a prisoner of war [65]. The *functoriality* of sheaves lends them particularly well to the type of principled discretization of geometric structures that we are interested in. In particular, we consider a Hilbert-space valued version of *cellular sheaves* on graphs, as introduced in [44]. Intuitively, they can be understood as generalized graph structures with signals valued in Hilbert spaces along with edge-wise coupling rules. For a more thorough introduction to cellular sheaves, see Appendix G.5.

In this work, our primary interest is in Hilbert sheaves that represent discretizations of the structure of a Hilbert bundle over a manifold. In particular, we desire a spatial discretization such that we can recover an appropriate discrete analogueof the Hilbert bundle's connection Laplacian. Formally, given an iid random sample $\mathcal{X}_n \subset \mathcal{M}$ from the uniform distribution (see Def. 23), we have the following.

Definition 3 (Hilbert Cellular Sheaf from a Hilbert Bundle)**.** *For a given Hilbert bundle $(\mathcal{M}, \mathcal{E}, \nabla)$ with sampled points $\mathcal{X}_n = \{x_1, \dots, x_n\} \subset \mathcal{M}$, fix a geodesic γ_{ij} between x_i and x_j, for all $i < j$. Further, let $m_{\gamma_{ij}}$ denote the midpoint of this geodesic. Consider the graph $G_n = (\mathcal{X}_n, E)$ with an undirected edge e_{ij} between x_i and x_j, for each $i < j$. The associated **Hilbert cellular sheaf** $\mathcal{F}_n^t$ on G_n with bandwidth parameter t is given by the following assignments:*

- *The Hilbert space $\mathcal{F}_n^t(x_i) := \mathcal{E}_{x_i}$ for each $x_i \in \mathcal{X}_n$, referred to as the node stalk over $x_i \in \mathcal{X}_n$.*
- *The Hilbert space $\mathcal{F}_n^t(e_{ij}) := \mathcal{E}_{m_{\gamma_{ij}}}$ for each $e_{ij} \in E$, referred to as the edge stalk over $e_{i,j} \in E$.*
- *For each edge $e_{ij} \in E$ with bounding vertices x_i, x_j, a pair of bounded linear restriction maps*

$$(\mathcal{F}_n^t)_{x_i \le e_{ij}} := \sqrt{k_{ij}^t}\, P_{x_i \to m_{\gamma_{ij}}} : \mathcal{F}_n^t(x_i) \to \mathcal{F}_n^t(e_{ij}),$$
$$(\mathcal{F}_n^t)_{x_j \le e_{ij}} := \sqrt{k_{ij}^t}\, P_{x_j \to m_{\gamma_{ij}}} : \mathcal{F}_n^t(x_j) \to \mathcal{F}_n^t(e_{ij}), \tag{3}$$

where $k_{ij}^t = e^{-d_{\mathcal{M}}(x_i, x_j)^2/4t}$, with $d_{\mathcal{M}}$ the geodesic distance on $\mathcal{M}$, and $P_{x_i \to m_{\gamma_{ij}}}$ denotes the unitary parallel transport map on $\mathcal{E}$ between x_i and $m_{\gamma_{ij}}$.

For the sake of exposition, the choice of sample and corresponding geodesic paths will often be suppressed, so our parallel transports will be denoted as $P_{x_i \to m_{ij}}$. Also, note that for $n < m$ and $\mathcal{X}_n \subset \mathcal{X}_m$, we assume each additional point is again sampled iid from the uniform distribution on $\mathcal{M}$. For the categorically-minded reader, we remark that our sheaf is constructed such that refining our sample then leads to a *subfunctor* $\mathcal{F}_n^t \subset \mathcal{F}_m^t$. For the Hilbert sheaf $\mathcal{F}_n^t$ on the graph $G_n = (\mathcal{X}_n, E)$, a *signal* is an element of the Hilbert space

$$C^0(\mathcal{F}_n^t; G_n) := \bigoplus_{x_i \in \mathcal{X}_n} \mathcal{F}_n^t(x_i). \tag{4}$$

Example 4. *If $\mathcal{F}_n^t$ encodes univariate spatiotemporal data, each node stalk can be chosen as $\mathcal{F}_n^t(x_i) = L^2(\mathbb{R})$. Then $C^0(\mathcal{F}_n^t; G_n) = \bigoplus_{x_i \in \mathcal{X}_n} L^2(\mathbb{R})$, so a signal assigns a full time series to every node, recovering the usual notion of a node-time graph signal.*

Example 5. *In one dimension, a probability distribution $\mu \in \mathcal{P}_2(\mathbb{R})$ can be represented by its quantile function $Q_\mu \in L^2([0,1])$, and the Wasserstein distance becomes the L^2 distance between*

quantiles [99]. Thus, by choosing node stalks $\mathcal{F}_n^t(x_i) = L^2([0,1])$, a signal assigns a full probability distribution to each graph node, recovering the distributional graph-signal setting of [57, 112].

Finally, analogous to the construction of the connection Laplacian Δ_∇, we may construct the Hilbert sheaf Laplacian. Further details, such as self-adjointness, are discussed in Appendix G.5.

Definition 4 (Hilbert Sheaf Laplacian)**.** *Let $\mathcal{F}_n^t$ be the Hilbert sheaf on the graph $G_n = (\mathcal{X}_n, E)$ induced by Def, 3. Fix an orientation for each edge $e \in E$. The **Hilbert sheaf Laplacian** is the bounded linear operator*

$$\Delta_{\mathcal{F}_n^t} : C^0(\mathcal{F}_n^t; G_n) \to C^0(\mathcal{F}_n^t; G_n) \tag{5}$$

defined, for a signal $S \in C^0(\mathcal{F}_n^t; G_n)$ and at a node $x_i \in \mathcal{X}_n$, by

$$(\Delta_{\mathcal{F}_n^t} S)_{x_i} = \sum_{\substack{e \in E: \\ e = \{x_i, x_j\}}} (\mathcal{F}_n^t)^*_{x_i \leq e} \left((\mathcal{F}_n^t)_{x_i \leq e} S_{x_i} - (\mathcal{F}_n^t)_{x_j \leq e} S_{x_j} \right), \tag{6}$$

where x_j denotes the other endpoint of e, and $(\mathcal{F}_n^t)^*_{x_i \leq e}$ is the adjoint of the restriction map $(\mathcal{F}_n^t)_{x_i \leq e}$.

Intuitively, $\Delta_{\mathcal{F}_n^t}$ measures how much a signal fails to be locally consistent across edges: before comparing S_{x_i} and S_{x_j}, both values are mapped into the common edge stalk $\mathcal{F}_n^t(e)$ by the restriction maps. Thus, it is a broad generalization of a graph Laplacian, with restriction maps replacing scalar edge weights. The Hilbert sheaf Laplacian $\Delta_{\mathcal{F}_n^t}$ is a self-adjoint bounded linear operator. Once the manifold is sampled and the induced sheaf Laplacian is computed, space-discretized HilbNets, which are still not implementable due to the infinite-dimensional signals, are simply given by Def, 2 with the connection Laplacian of $\mathcal{E}$ replaced by the sheaf Laplacian of $\mathcal{F}_n^t$, i.e., by $\Omega(\mathcal{F}_n^t, \Delta_{\mathcal{F}_n^t}, \mathcal{W}, \sigma)$.

Signal Sampling. Hilbert cellular sheaves are the structures that arise when we sample our base manifold but faithfully record the potentially infinite-dimensional signal in each fiber. In practice, we typically only have access to a sampled or compressed version of the signal as well. For instance, when considering a timeseries $S \in L^2(\mathbb{R})$, we may use the orthogonal Fourier basis $\overline{\{e^{ik\theta}\}}_{k \in \mathbb{Z}} = L^2(\mathbb{R})$ and then record a compressed representation with respect to this basis i.e. $[\langle S, e^{-id\theta} \rangle, \ldots, \langle S, e^{id\theta} \rangle]$ for some d. We can consider fiber-wise orthogonal projections with respect to any chosen basis in the Hilbert bundle setting as a principled approach to discretizing Hilbert bundle signals.

Proposition 1. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle, with strictly infinite-dimensional generic Hilbert-space fiber $\mathcal{H}$. Fix an orthogonal basis $\mathcal{B} = \{e_1, e_2, ...\}$ of $\mathcal{H}$ and let $\mathcal{H}_d := \mathrm{span}(e_1, e_2, ..., e_d)$. Then there exists a smooth map of bundles*

$$\Pi_d : \mathcal{E} \to \mathcal{E}_d \tag{7}$$

where $\mathcal{E}_d$ is a d-dimensional vector bundle with generic fiber $\mathcal{H}_d$ and at each $x \in \mathcal{M}$, $\Pi_d|_{\mathcal{E}_x} : \mathcal{E}_x \to \mathcal{E}_{d,x}$ recovers the usual orthogonal projection map. See Appendix H.4 for details.

Applying Proposition 1 to $(\mathcal{M}, \mathcal{E}, \nabla)$, we obtain $(\mathcal{M}, \mathcal{E}_d, \nabla_d)$, to which we may apply the spatial discretization of Def.3 to construct the cellular sheaf $\mathcal{F}_{n,d}^t$ with d-dimensional stalks. We refer to $\mathcal{F}_{n,d}^t$ as a *network sheaf*. The signals on this sheaf are then sampled Hilbert bundle signals, i.e., we can discretize $S \in L^2(\mathcal{M}, \mathcal{E})$ as a nd-dimensional vector $\mathbf{s}_{n,d} := \Pi_d S \in C^0(\mathcal{F}_{n,d}^t; G_n) \subseteq \mathbb{R}^{nd}$, stacking the d-dimensional orthogonal projections over the n sampled locations, with respect to the chosen basis $\mathcal{B}$. In this case, the restriction maps can be written as matrices, thus the sheaf Laplacian becomes a block matrix $\Delta_{\mathcal{F}_{n,d}^t} \in \mathbb{R}^{nd \times nd}$ whose (i,j)-block maps the discretized stalk over x_j to the discretized stalk over x_i and is given by

$$(\Delta_{\mathcal{F}_{n,d}^t})_{ij} = \begin{cases} \sum_{r:\, e_{ir} \in E} k_{ir}^t I_d, & i = j, \\ -P_{x_j \to x_i}^{(d, e_{ij})}, & i \neq j \text{ and } e_{ij} \in E, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $P_{x_j \to x_i}^{(d,e_{ij})} := k_{ij}^t \left(P_{x_i \to m_{ij}}^{(d)} \right)^* P_{x_j \to m_{ij}}^{(d)}$, and $P_{x_j \to m_{ij}}^{(d)}$ denotes the restriction of the parallel transport map from $\mathcal{E}_{x_j}$ to $\mathcal{E}_{m_{ij}}$ to the corresponding d-dimensional subbundles in the image of Π_d. We may thus use this Laplacian to build an implementable sheaf convolutional architecture as follows.

Definition 5 ((n,d)-Hilbert bundle convolutional neural network)**.** *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle, with generic Hilbert-space fiber $\mathcal{H}$ and corresponding basis $\mathcal{B}$. A (n,d)-Hilbert bundle*

*convolutional neural network, or (n,d)-**HilbNet**, is specified by a filter bank $\mathcal{W} = \{g^\ell_{u,q}\}_{\ell,u,q}$ with $g^\ell_{u,q} \in L^\infty_c(\mathbb{R})$, a Lipschitz continuous nonlinear activation $\sigma : \mathbb{R} \to \mathbb{R}$, the choice of a d-dimensional subbasis of $\mathcal{B}$ and sample $\mathcal{X}_n \subset \mathcal{M}$. Given input sampled signals $\mathbf{s}_{n,d,1}, ..., \mathbf{s}_{n,d,F_0} \in C^0(\mathcal{F}^t_{n,d}; G_n)$, the L-layer network output is obtained by the recursion*

$$\mathbf{s}^{\ell+1}_{n,d,u} = \sigma\left(\sum_{q=1}^{F_\ell} g^\ell_{u,q}(\Delta_{\mathcal{F}^t_{n,d}})\mathbf{s}^\ell_{n,d,q}\right), \qquad \ell = 0, \dots, L-1, \qquad \mathbf{s}^0_{n,d,q} = \mathbf{s}_{n,d,q}. \tag{9}$$

Discretized HilbNets are fully implementable and can be compactly written using Def. 2 with the connection Laplacian of $\mathcal{E}$ replaced by the sheaf Laplacian of $\mathcal{F}^t_{n,d}$, i.e., by $\Omega(\mathcal{F}^t_{n,d}, \Delta_{\mathcal{F}^t_{n,d}}, \mathcal{W}, \sigma)$.

Example 6. *In the case that we consider our filter bank $\mathcal{W}$ to consist of order K polynomials, the (n,d)-HilbNet can be written as a novel variant of sheaf neural networks [50, 11] given by*

$$\mathbf{S}^{\ell+1}_{n,d} = \sigma\left(\sum_{k=0}^{K-1} (\Delta_{\mathcal{F}^t_{n,d}})^k \mathbf{S}^\ell_{n,d}\mathbf{W}_{\ell,k}\right) \in \mathbb{R}^{nd}. \tag{10}$$

where the matrices $\mathbf{S}^\ell_{n,d} \in \mathbb{R}^{nd\times F_\ell}$, and $\{\mathbf{W}_{\ell,k}\}_k$, with $\mathbf{W}_{\ell,k} \in \mathbb{R}^{F_\ell \times F_{\ell+1}}$ collect the sampled signals and the learnable filter weights at each layer, respectively.

5 Theoretical Convergence Guarantees

Our main result may be understood as a far-reaching generalization of the convergence result of Belkin & Niyogi [14]. Consider a random sample $\mathcal{X}_n \subset \mathcal{M}$ and the corresponding geometric graph $G_n = (\mathcal{X}_n, E)$. It is established in [14] that as sampling density increases, the weighted graph Laplacian converges to the manifold Laplace-Beltrami operator in probability. We analogously show that the Hilbert sheaf Laplacian $\Delta_{\mathcal{F}_n}$ over G_n converges to Δ_∇, thus recovering the results of [14] as the special case $\mathcal{E} = \mathcal{M} \times \mathbb{R}$. Our proof, presented in Appendix H, is inspired by the strategy of [14] but with the necessary non-trivial modifications to accommodate the simultaneous generalization to cellular sheaves instead of graphs and to infinite-dimensional Hilbert-spaces. In order to state our results, we require the following intermediary operator.

Definition 6. *(Point-Cloud Extension of Sheaf Laplacian) Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle and consider a sample $\mathcal{X}_n \subset \mathcal{M}$. Then the corresponding Hilbert sheaf Laplacian $\Delta_{\mathcal{F}^t_n}$ may be extended to the **point-cloud Laplacian** $\hat{\Delta}_{\mathcal{F}^t_n}$, an operator on $L^2(\mathcal{M}, \mathcal{E})$ via*

$$(\hat{\Delta}_{\mathcal{F}^t_n} S)(x) = \frac{1}{n}\sum_j e^{-d_{\mathcal{M}}(x,x_j)^2/4t}\big(S(x) - P_{x_j \to x} S(x_j)\big) \tag{11}$$

As such, we are able to consider the sheaf-level and bundle-level Laplacians as operators on the same space through this extension. In this setting, we then have the following convergence result.

Theorem 1. *(Convergence of Hilbert Sheaf Laplacian) Let $\mathcal{M}$ be a m-dimensional closed Riemannian manifold. Further, let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle and associated connection Laplacian Δ_∇. Fix a section $S \in C^3(\mathcal{M}, \mathcal{E})$. Consider a random sample $\mathcal{X}_n = \{x_1, x_2, \cdots, x_n\} \subset \mathcal{M}$. Let $\mathcal{F}^t_n$ be the induced Hilbert cellular sheaf with bandwidth t. Then we have, for any $x \in \mathcal{M}$,*

$$\lim_{n\to\infty} \frac{1}{t_n\,(4\pi t_n)^{\frac{m}{2}}} \hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) = \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S(x) \quad \textit{in probability,} \tag{A}$$

with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$. Further, if $S \in C^4(\mathcal{M}, \mathcal{E})$, we have

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X}}\left[\left\| \frac{1}{t_n\,(4\pi t_n)^{\frac{m}{2}}} \hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S(x)\right\|^2_{L^2}\right] = 0 \quad \textit{in } L^2\textit{-norm.} \tag{B}$$

Our framework may be seen as concurrently generalizing the convergence results for the weighted graph Laplacian of [14] as well as the graph connection Laplacian of [93] to allow for arbitrary bundles, with potentially infinite-dimensional fibers, equipped with an arbitrary choice of connection. Such convergence results have previously served as the basis of transferability and robustness results in geometric deep learning [104, 103, 67], as well as to justify the development of numerous Laplacian-based manifold learning techniques [13, 28, 108, 97, 85]. We may likewise develop generalizations of these results for implementable sheaf Laplacians by discretizing in the signal-domain.

Finite Rank Convergence. Consider a signal S that has been sampled to $\mathbf{s}_{n,d}$ as per Proposition 1. We then have the following theoretical guarantee, which formalizes the intuitive notion that the fully discretized sheaf Laplacian converges to true connection Laplacian as we take an increasingly refined sample of both the underlying manifold and the signal.

Theorem 2 (Finite-Rank Approximation)**.** *Consider the setting of Theorem 1 with a section $S \in C^4(\mathcal{M}, \mathcal{E})$, for $\mathcal{E}$ a strictly infinite-dimensional Hilbert bundle. Then there exists a sequence of finite rank approximating sheaves $\mathcal{F}^{t_n}_{n,d_n}$ such that*

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X}} \left[\left\| \frac{1}{t_n(4\pi t_n)^{m/2}} \hat{\Delta}_{\mathcal{F}^{t_n}_{n,d_n}} \mathbf{s}_{n,d_n} - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S \right\|^2_{L^2} \right] = 0, \tag{12}$$

with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$.

While this result, as described in Appendix B, can pave the way for new developments of Laplacian-based manifold learning, we here restrict our focus to its consequences for (n, d)-HilbNets.

Corollary 2.1 (Convergence in Architecture)**.** *Under the hypotheses of Theorem 2, let $\{d_n\}_n$ be the required sequence. Fix a fiber-wise nonlinearity σ that is C_σ-Lipschitz in the corresponding fiber norms and choice of filter bank $\mathcal{W}$. Then, the output of the discrete (n, d)-HilbNet converges to the output of the continuous HilbNet architecture in the sense that*

$$\Omega(\mathcal{F}^{t_n}_{n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{n,d_n}}, \mathcal{W}, \sigma) \to \Omega(\mathcal{E}, \Delta_\nabla, \mathcal{W}, \sigma) \quad \textit{in mean squared error}, \tag{13}$$

as the sampling density $n, d_n \to \infty$.

Corollary 2.2. *(Transferability) Let $\{\mathcal{X}_n\}_{n=1}^\infty$ and $\{\mathcal{Y}_n\}_{n=1}^\infty$ be independent sequences of random samples of $\mathcal{M}$. Let $\{d_n\}_{n=1}^\infty$ be a sequence such that the conclusion of Theorem 2 holds for both samplings. For any fiber-wise nonlinearity σ that is C_σ-Lipschitz in the corresponding fiber norms and any filter bank $\mathcal{K}$, we then have that,*

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X},\mathcal{Y}} \left[\left\| \Omega(\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}}, \mathcal{W}, \sigma) - \Omega(\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}}, \mathcal{W}, \sigma) \right\|^2_{L^2} \right] = 0. \tag{14}$$

Further, one may derive a sample-independent quantitative bound for the L^2 disagreement $\|\Omega(\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}}, \mathcal{W}, \sigma) - \Omega(\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}}, \mathcal{W}, \sigma)\|_{L^2}$. See Appendix H for details.

By these results, we may understand the (n, d)-HilbNets as the principled discretization of continuous HilbNets. These may also be understood as robustness results for (n, d)-HilbNets, as they establish that the architecture is scale-consistent and is stable against resampling of the base manifold. Notably, by the generality of HilbNets, most existing geometric convolutional architectures can be understood as instances of HilbNets for a particular choice of bundle and connection. As such, our results may also be seen as extending the transferability guarantees of [104, 103, 67] to a larger class of architectures and data modalities. See Appendix C for further discussion.

	Free $O(m)$		Circulant		Frozen identity (GCN)	
n	Det.	Theory	Det.	Theory	Det.	Theory
16	$(1.71\pm0.17)\times10^{-7}$	0	$(1.84\pm0.29)\times10^{-2}$	1.84×10^{-2}	$(2.30\pm0.31)\times10^{-2}$	2.30×10^{-2}
32	$(1.28\pm0.57)\times10^{-7}$	0	$(1.18\pm0.14)\times10^{-2}$	1.18×10^{-2}	$(1.46\pm0.18)\times10^{-2}$	1.46×10^{-2}
64	$(1.82\pm0.47)\times10^{-7}$	0	$(1.03\pm0.14)\times10^{-2}$	1.03×10^{-2}	$(1.29\pm0.16)\times10^{-2}$	1.29×10^{-2}
128	$(2.22\pm0.25)\times10^{-7}$	0	$(8.94\pm0.71)\times10^{-3}$	8.93×10^{-3}	$(1.11\pm0.09)\times10^{-2}$	1.11×10^{-2}
256	$(1.87\pm0.35)\times10^{-7}$	0	$(7.80\pm0.12)\times10^{-3}$	7.80×10^{-3}	$(9.67\pm0.18)\times10^{-3}$	9.67×10^{-3}

Table 1: Synthetic transport recovery. **Det.**: edge-MSE of the learned transports, bounded below by Theory (Appendix F.1). **Theory**: analytical squared Frobenius projection distance of $\tilde{P}^{LC}_{x_j\to x_i}$ onto the variant's hypothesis class.

6 Experimental Results

A key practical advantage of the (n, d)-HilbNets architecture in comparison to existing approaches for processing graph signals is our use of parallel transport, which in practice can be known or learned. For instance, these transport operators allow us to incorporate principled signal-level geometric priors in concert with the spatial priors of existing spatiotemporal GCNs. This is well-aligned with the thesis of geometric deep learning that the principled incorporation of geometric priors improves

Model	Params	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
FC-LSTM [69]	>1.5M	3.44	6.30	9.6	3.77	7.23	10.9	4.37	8.69	13.2
STAEformer [71]	∼1.3M	2.65	5.11	6.85	2.97	6.00	8.13	3.34	7.02	9.70
MLP fiber baseline	5,212	3.131±0.004	6.074±0.007	8.271±0.021	3.775±0.005	7.496±0.013	10.626±0.046	4.690±0.011	9.184±0.014	14.341±0.097
Spatiotemporal graph baseline	8,908	3.453±0.080	6.709±0.220	9.241±0.279	4.160±0.117	8.158±0.184	11.916±0.491	5.277±0.093	10.102±0.128	16.006±0.662
HilbNet, frozen identity (GCN)	5,756	3.092±0.007	5.920±0.013	8.218±0.065	3.713±0.010	7.312±0.020	10.520±0.061	4.608±0.034	8.991±0.043	14.166±0.240
HilbNet, circulant	11,656	2.939±0.021	5.630±0.061	7.908±0.067	3.409±0.032	6.765±0.092	9.844±0.125	4.059±0.049	8.149±0.114	12.471±0.195
HilbNet, free $O(T)$	119,036	**2.923±0.013**	**5.586±0.048**	**7.808±0.083**	**3.372±0.023**	**6.732±0.066**	**9.507±0.096**	**3.938±0.030**	**8.042±0.101**	**11.642±0.136**

Model	Params	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
FC-LSTM [69]	>1.5M	2.05	4.19	4.8	2.20	4.55	5.2	2.37	4.96	5.7
STAEformer [71]	∼1.4M	1.31	2.78	2.76	1.62	3.68	3.62	1.88	4.34	4.41
MLP fiber baseline	5,212	1.459±0.003	3.145±0.017	3.026±0.020	1.942±0.004	4.378±0.017	4.385±0.045	2.513±0.004	5.658±0.015	6.209±0.038
Spatiotemporal graph baseline	8,908	**1.400±0.002**	2.980±0.016	**2.924±0.003**	1.850±0.002	4.162±0.009	4.222±0.009	2.388±0.004	5.395±0.023	5.924±0.032
HilbNet, frozen identity (GCN)	5,756	1.439±0.003	3.077±0.022	2.985±0.007	1.901±0.006	4.264±0.021	4.319±0.015	2.446±0.007	5.495±0.019	6.020±0.032
HilbNet, circulant	15,366	1.413±0.002	2.971±0.018	2.982±0.015	1.806±0.003	**3.950±0.015**	4.162±0.037	2.211±0.014	**4.866±0.048**	5.386±0.047
HilbNet, free $O(T)$	190,268	1.417±0.002	**2.969±0.015**	3.058±0.013	**1.793±0.005**	3.958±0.026	**4.127±0.037**	**2.181±0.003**	4.873±0.019	**5.214±0.045**

Table 2: METR-LA (top) and PEMS (bottom) traffic forecasting results. Bottom block of each table: our experiments (mean ± standard deviation over five seeds). Top block of each table: external baselines reported as in the cited papers (parameter counts estimated from their published configurations). MAPE is in percent. Lower is better for all metrics.

performance, particularly in the low-data or small-model regimes. For further discussion on strategies for either hand-crafting or learning parallel transport operators, see Appendix D. Here, we first validate our setup for a synthetic dataset realized from discretizing a known Hilbert bundle in information geometry, and then consider performance on real-world spatiotemporal graph benchmarks based upon traffic forecasting. In all the experiments, we use the polynomial (n, d)-HilbNet from (10) and learned transport maps. Transports are parametrized as free $O(d)$ maps (products of R Householder reflections), orthogonal circulant maps, or frozen identities (see Appendix E for details). For traffic forecasting, each HilbNet variant stacks $L{=}2$ sheaf convolutional layers with polynomial filters of order $K{=}2$ and widths 16 and 32, followed by a per-node linear readout over the full horizon. Full architectural details and hyperparameters are in Appendix F (Tables 4 and 5). Code is available at `https://github.com/clabat9/hilbnet`.

Synthetic Experiments. We first consider a task where, for a known Hilbert bundle and connection, we train a discretized HilbNet to predict the true parallel transport operators. Following [72], the base manifold is $\mathcal{M} = \mathrm{Sym}^{++}(p)$ equipped with the Otto-Wasserstein metric, each $\Sigma \in \mathcal{M}$ parameterizing a density $\mathcal{N}(0, \Sigma)$. The ambient fiber $\mathcal{H}_\Sigma = L^2(\rho_\Sigma; \mathbb{R}^p)$ is genuinely infinite-dimensional; the computational fiber is the Otto-velocity image of covariance perturbations, a sub-bundle whose fibers are already finite-dimensional with $d = p(p+1)/2$, and on which the Levi-Civita transports $P^{LC}_{x_i \to m_{ij}}$ admit a closed-form. We sample n points, build a kNN graph G_n ($k{=}8$) under W_2, and assemble the network sheaf $\mathcal{F}^t_{n,d}$ and its Laplacian $\Delta_{\mathcal{F}^t_{n,d}} \in \mathbb{R}^{nd \times nd}$ from Def, 3. We consider three transport parametrizations from Appendix D (averaged over 3 seeds): free $O(d)$ (Householder), circulant, and frozen identity (a usual GCN [59]), to recover the Levi-Civita transports in Cholesky-rescaled coordinates. As the reader can notice in Table 1, the free class recovers $\tilde{P}^{LC}_{x_j \to x_i}$ to numerical precision $\approx 2 \cdot 10^{-7}$, while each restricted class converges to its analytical Frobenius-projection plateau to within 0.05%. This confirms that the transport hypothesis class constrains the per-edge restriction maps of $\Delta_{\mathcal{F}^t_{n,d}}$ in a quantitatively predictable way. More experiments and details are in Appendix F.1

Traffic Forecasting. We evaluate HilbNets on real-world spatiotemporal traffic-speed forecasting, where each node of a road-network graph carries a time-series fiber with $d = T$ observed time steps. This is a natural instance of the Hilbert-bundle framework: the base graph encodes spatial proximity among sensors, while the edgewise transports $P^{e_{ij}}_{j \to i}$ from Appendix D control how temporal fibers are aligned before filtering. We test on two standard benchmarks, METR-LA [69] and PEMS-BAY [69], predicting future speeds at horizons 3, 6, and 12. We compare the same three HilbNet transport classes, frozen identity (a GCN), circulant, and free $O(T)$, against a fiber-only MLP and a spatiotemporal graph baseline obtained by stacking GCN layers with one-dimensional convolutional layers processing the temporal dimension, all sharing the same polynomial sheaf filter order, readout, and forecasting loss. Full experimental details are given in Appendix F.2. Table 2 reports MAE, RMSE, and MAPE (mean ± std over five seeds). On both datasets, learning non-trivial transports consistently improves over frozen identity at all horizons, confirming that the sheaf structure helps beyond the usual graph structure. The free $O(T)$ class achieves the best overall accuracy, but the circulant variant is competitive while using roughly one tenth of the transport parameters, supporting the use of structured, physically motivated transport priors for spatiotemporal data. This confirms that geometric inductive biases can lead to better performance in low-data regimes or comparable performance in normal regimes with substantially fewer parameters.

7 Conclusions

We introduced a novel convolutional learning framework for infinite-dimensional signals over a manifold using Hilbert bundles, a setting that concurrently unifies and generalizes existing approaches. It allows us to consider arbitrary connection Laplacians, a more general class of filters via the Borel calculus, and thus applications of the resulting filters to potentially infinite-dimensional signals. We defined HilbNets as stacks of Hilbert bundle filters and pointwise non-linearity. We consequently introduced a practically implementable *(n,d)-HilbNet* via the theory of *Hilbert cellular sheaves*, and proved that this discretized architecture converges to the continuous architecture in the limit. Notably, our convergence in architecture is derived from a novel extension of the Laplacian convergence result of [14] to the setting of Hilbert sheaves and Hilbert bundles, and we believe this result will be of independent interest to the broader machine learning community. Lastly, we verified the benefits of integrating domain-specific geometric priors through experiments with discretized HilbNets on synthetic and real-world data. Overall, we envision the prospective impact of our contributions as two-fold: the HilbNet framework allows for the principled development of domain-specific architectures through appropriate choices of connection, filter bank, and manifold- and signal-alignment measures, while our Hilbert Laplacian convergence theorem lays the theoretical groundwork for the development of Laplacian-based manifold-theoretic techniques in the setting of infinite-dimensional signals, spanning from mechanistic interpretability to self-supervised learning methods. A more detailed discussion on impact and future directions is presented in Appendix B.

Limitations. Our theoretical guarantees rest on some assumptions that may fail to hold in practice, as is usually the case. The convergence results (Theorems 1–2) require the base manifold $\mathcal{M}$ to be closed (compact without boundary), sections to be sufficiently smooth, and samples to be drawn i.i.d. from the uniform distribution. Real-world sensor networks, such as those in our traffic experiments, are neither uniformly sampled nor necessarily supported on compact manifolds. Moreover, measured signals are typically noisy rather than smooth. However, many works, including the foundational results of [14], as well as related works on vector diffusion maps [92, 93] and manifold-based learning [28], assume compact manifolds and uniform or smooth sampling densities, yet are routinely and successfully applied under weaker conditions. We expect that HilbNets would behave similarly and, in fact, our numerical results confirm that HilbNets likewise remain effective under these standard approximations, generally outperforming matched baselines that lack the principled bundle-geometric structure. On the computational side, the network sheaf Laplacian scales quadratically in the product of spatial and fiber dimensions, which may become prohibitive for very large graphs or high-dimensional signal discretizations without further sparsification or approximation strategies. In practice, this cost can be contained by building the sheaf on a sparse kNN graph rather than on a dense sample, and by structured, parameter-efficient transport parametrizations such as circulant maps (Appendix E). Finally, broader and tailored empirical validation on other infinite-dimensional signal types, such as functional data on manifolds, remains an important direction for future work.

Acknowledgments and Disclosure of Funding

C.B. and F.D. acknowledge support from the National Institutes of Health Grant 1R01ES037156-01.The authors declare no competing interests.

References

[1] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4277–4280, 2012. doi: 10.1109/ICASSP.2012.6288864.

[2] Manasvi Aggarwal and Musti Narasimha Murty. *Machine Learning in Social Networks: Embedding Nodes, Edges, Communities, and Graphs*. SpringerBriefs in Applied Sciences and Technology. Springer, Singapore, 2021. doi: 10.1007/978-981-33-4022-0.

[3] Antonio Ambrosetti and Giovanni Prodi. *A Primer of Nonlinear Analysis*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, UK, 1995. ISBN 9780521485739.

[4] Scott Axelrod, Steve Della Pietra, and Edward Witten. Geometric quantization of Chern–Simons gauge theory. *J.Differential Geom.*, 33(3):787–902, 1991. doi: 10.4310/jdg/1214446565.

[5] Jacob Bamberger, Federico Barbero, Xiaowen Dong, and Michael M. Bronstein. Bundle neural network for message diffusion on graphs. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=scI9307PLG.

[6] Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Trans. on Signal Processing*, 68:2992–3007, 2020. doi: 10.1109/TSP.2020.2981920.

[7] Federico Barbero, Cristian Bodnar, Haitz Sáez de Ocáriz Borde, Michael Bronstein, Petar Veličković, and Pietro Liò. Sheaf neural networks with connection Laplacians. In *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*, volume 196 of *Proceedings of Machine Learning Research*, pp. 28–36. PMLR, 2022. URL https://proceedings.mlr.press/v196/barbero22a.html.

[8] Thomas Batard. Heat equations on vector bundles—application to color image regularization. *Journal of Mathematical Imaging and Vision*, 41(1):59–85, 2011. doi: 10.1007/s10851-011-0265-3.

[9] Claudio Battiloro, Zhiyang Wang, Hans Riess, Paolo Di Lorenzo, and Alejandro Ribeiro. Tangent bundle filters and neural networks: From manifolds to cellular sheaves and back. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023. doi: 10.1109/ICASSP49357.2023.10096934.

[10] Claudio Battiloro, Lucia Testa, Lorenzo Giusti, Stefania Sardellitti, Paolo Di Lorenzo, and Sergio Barbarossa. Generalized simplicial attention neural networks. *IEEE Transactions on Signal and Information Processing over Networks*, 10:833–850, 2024. doi: 10.1109/TSIPN.2024.3485473.

[11] Claudio Battiloro, Zhiyang Wang, Hans Riess, Paolo Di Lorenzo, and Alejandro Ribeiro. Tangent bundle convolutional learning: From manifolds to cellular sheaves and back. *IEEE Transactions on Signal Processing*, 72:1892–1909, 2024. doi: 10.1109/TSP.2024.3379862.

[12] M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005. doi: 10.1023/B:VISI.0000043755.93987.aa.

[13] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, volume 14, pp. 585–591. MIT Press, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/file/f106b7f99d2cb30c3db1c3cc0fde9ccb-Paper.pdf.

[14] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008. ISSN 0022-0000. doi: 10.1016/j.jcss.2007.08.006. URL https://www.sciencedirect.com/science/article/pii/S0022000007001274. Learning Theory 2005.

[15] Nicole Berline, Ezra Getzler, and Michèle Vergne. *Heat Kernels and Dirac Operators*. Grundlehren Text Editions. Springer, Berlin, Heidelberg, 2004. doi: 10.1007/978-3-642-58088-8.

[16] Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yu Guang Wang, Pietro Liò, Guido Montúfar, and Michael M. Bronstein. Weisfeiler and Lehman go cellular: CW networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2625–2640. Curran Associates, Inc., 2021.

[17] Cristian Bodnar, Fabrizio Frasca, Yu Guang Wang, Nina Otter, Guido Montúfar, Pietro Liò, and Michael Bronstein. Weisfeiler and Lehman go topological: Message passing simplicial networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1026–1037, 2021.

[18] Cristian Bodnar, Francesco Di Giovanni, Benjamin Chamberlain, Pietro Liò, and Michael Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18527–18541. Curran Associates, Inc., 2022. doi: 10.52202/068431-1346.

[19] Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar. Spherical Fourier neural operators: Learning stable dynamics on the sphere. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2806–2823. PMLR, 2023.

[20] Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Peter Deisenroth. Matérn Gaussian processes on Riemannian manifolds. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12426–12437, 2020.

[21] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. doi: 10.1109/MSP.2017.2693418.

[22] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[23] Jochen Brüning and Matthias Lesch. Hilbert complexes. *Journal of Functional Analysis*, 108 (1):88–132, 1992. doi: 10.1016/0022-1236(92)90147-B.

[24] Alessia Caponera and Domenico Marinucci. Asymptotics for spherical functional autoregressions. *The Annals of Statistics*, 49(1):346–369, 2021. doi: 10.1214/20-AOS1959.

[25] Henri Cartan. *Differential Calculus*. Hermann, Paris, 1971. ISBN 0-395-12033-0.

[26] Gengxiang Chen, Xu Liu, Qinglu Meng, Lu Chen, Changqing Liu, and Yingguang Li. Learning neural operators on Riemannian manifolds. *National Science Open*, 3(6), 2024. doi: 10.1360/nso/20240001.

[27] Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[28] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. doi: 10.1016/j.acha.2006.04.006.

[29] Justin Michael Curry. *Sheaves, cosheaves and applications*. PhD thesis, University of Pennsylvania, 2014.

[30] Gabriele D'Acunto and Claudio Battiloro. The relativity of causal knowledge. In *The 41st Conference on Uncertainty in Artificial Intelligence*, 2025. URL `https://openreview.net/forum?id=aS8mPNs5u5`.

[31] Xiongtao Dai and Hans-Georg Müller. Principal component analysis for functional data on Riemannian manifolds and spheres. *The Annals of Statistics*, 46(6B):3334–3361, 2018. doi: 10.1214/17-AOS1660.

[32] Rumen Dangovski, Li Jing, Charlotte Loh, Seungwook Han, Akash Srivastava, Brian Cheung, Pulkit Agrawal, and Marin Soljačić. Equivariant self-supervised learning: Encouraging equivariance in representations. In *International Conference on Learning Representations*, 2022.

[33] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

[34] Pim De Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. In *International Conference on Learning Representations*, 2021.

[35] Leonardo Di Nino, Gabriele D'Acunto, Sergio Barbarossa, and Paolo Di Lorenzo. Learning the structure of connection graphs. In *ICASSP 2026 – 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 76–80, 2026.

[36] Manfredo P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser, Boston, 1992. ISBN 978-0817634902.

[37] Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, and Pietro Liò. Sheaf hypergraph networks. *Advances in Neural Information Processing Systems*, 36:12087–12099, 2023.

[38] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[39] Stefano Fiorini, Hakan Aktas, Iulia Duta, Pietro Morerio, Alessio Del Bue, Pietro Liò, and Stefano Coniglio. Sheaves reloaded: A direction awakening. In *The Fourteenth International Conference on Learning Representations*, 2026.

[40] Fernando Gama, Elvin Isufi, Geert Leus, and Alejandro Ribeiro. Graphs, convolutions, and neural networks: From graph filters to graph neural networks. *IEEE Signal Processing Magazine*, 37(6):128–138, 2020. doi: 10.1109/MSP.2020.3016143.

[41] Fernando Gama, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro. Convolutional neural network architectures for signals supported on graphs. *IEEE Transactions on Signal Processing*, 67(4):1034–1049, 2019.

[42] Robert Ghrist and Hans Riess. Cellular sheaves of lattices and the Tarski Laplacian. *Homology, Homotopy and Applications*, 24(1):325–345, 2022.

[43] Lorenzo Giusti, Claudio Battiloro, Lucia Testa, Paolo Di Lorenzo, Stefania Sardellitti, and Sergio Barbarossa. Cell attention networks. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2023.

[44] Julian J. Gould. *Cellular Sheaves of Hilbert Spaces*. PhD thesis, University of Pennsylvania, 2025.

[45] Francesco Grassi, Andreas Loukas, Nathanaël Perraudin, and Benjamin Ricaud. A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs. *IEEE Transactions on Signal Processing*, 66(3):817–829, 2018.

[46] Alexander Grigor'yan. *Heat Kernel and Analysis on Manifolds*, volume 47 of *AMS/IP Studies in Advanced Mathematics*. American Mathematical Society, Providence, RI, 2009.

[47] Enrico Grimaldi, Mario Edoardo Pandolfo, Gabriele D'Acunto, Sergio Barbarossa, and Paolo Di Lorenzo. Learning network sheaves for AI-native semantic communication. In *2025 59th Asilomar Conference on Signals, Systems, and Computers*, pp. 1692–1696, 2025.

[48] Alexandre Grothendieck. A general theory of fibre spaces with structure sheaf. National Science Foundation Research Project on Geometry of Function Space, Report 4, University of Kansas, Department of Mathematics, Lawrence, Kansas, 1955.

[49] Tyler Hanks, Hans Riess, Samuel Cohen, Trevor Gross, Matthew Hale, and James Fairbanks. Distributed multi-agent coordination over cellular sheaves. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pp. 3057–3064, 2025.

[50] Jakob Hansen and Thomas Gebhart. Sheaf neural networks, 2020. URL `https://arxiv.org/abs/2012.06333`.

[51] Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, Dec 2019. ISSN 2367-1734. doi: 10.1007/s41468-019-00038-7. URL `https://doi.org/10.1007/s41468-019-00038-7`.

[52] Jakob Hansen and Robert Ghrist. Learning sheaf Laplacians from smooth signals. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5446–5450, 2019. doi: 10.1109/ICASSP.2019.8683709.

[53] Jakob Hansen and Robert Ghrist. Opinion dynamics on discourse sheaves. *SIAM Journal on Applied Mathematics*, 81(5):2033–2060, 2021. doi: 10.1137/20M1341088.

[54] Edward J Hu, Moksh Jain, Eric Elmoznino, Younesse Kaddar, Guillaume Lajoie, Yoshua Bengio, and Esmeralda S. Whitammer. Amortizing intractable inference in large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Ouj6p4ca60.

[55] Hai Huang, Yann LeCun, and Randall Balestriero. Semantic tube prediction: Beating LLM data efficiency with JEPA. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, pp. 46752–46773, 2026.

[56] Michael Hutchinson, Alexander Terenin, Viacheslav Borovitskiy, So Takao, Yee Whye Teh, and Marc Deisenroth. Vector-valued Gaussian processes on Riemannian manifolds via gauge independent projected kernels. *Advances in Neural Information Processing Systems*, 34: 17160–17169, 2021.

[57] Feng Ji, Yanan Zhao, See Hian Lee, Kai Zhao, Wee Peng Tay, and Jielong Yang. Graph distributional signals for regularization in graph neural networks. *IEEE Transactions on Signal and Information Processing over Networks*, 11:670–682, Jul 2025. doi: 10.1109/tsipn.2025.3587400.

[58] Anran Jiao, Qile Yan, John Harlim, and Lu Lu. Solving forward and inverse partial differential equation problems on unknown manifolds via physics-informed neural operators. *SIAM Journal on Scientific Computing*, 48(1):C136–C163, 2026.

[59] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *Proc. of the 5th International Conference on Learning Representations (ICLR)*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

[60] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

[61] Nicolaas H. Kuiper. The homotopy type of the unitary group of Hilbert space. *Topology*, 3(1): 19–30, 1965. doi: 10.1016/0040-9383(65)90067-4.

[62] Serge Lang. *Differential and Riemannian Manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer, New York, NY, third edition, 1995. ISBN 978-0-387-94338-1. doi: 10.1007/978-1-4612-4182-9.

[63] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 86(11):2278–2324, 1998.

[64] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, Berlin, 1991.

[65] Jean Leray. L'anneau d'homologie d'une représentation. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 222:1366–1368, 1946.

[66] Geert Leus, Antonio G Marques, José MF Moura, Antonio Ortega, and David I Shuman. Graph signal processing: History, development, impact, and outlook. *IEEE Signal Processing Magazine*, 40(4):49–60, 2023.

[67] Ron Levie, Wei Huang, Lorenzo Bucci, Michael Bronstein, and Gitta Kutyniok. Transferability of spectral graph convolutional neural networks. *Journal of Machine Learning Research*, 22 (272):1–59, 2021.

[68] Daiyuan Li, Shreya Arya, and Robert Ghrist. Learning from frustration: Torsor CNNs on graphs. In *Proceedings of the 4th (2025) and 3rd (2024) NeurIPS Workshops on Symmetry and Geometry in Neural Representations*, volume 282 of *Proceedings of Machine Learning Research*, pp. 298–315, 2026.

[69] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR)*, 2018. URL https://openreview.net/forum?id=SJiHXGWAZ.

[70] Eardi Lila, John AD Aston, and Laura M Sangalli. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics*, 10(4):1854–1879, 2016.

[71] Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, and Xuan Song. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023. doi: 10.1145/3583780.3615160.

[72] Luigi Malagò, Luigi Montrucchio, and Giovanni Pistone. Wasserstein Riemannian geometry of Gaussian densities. *Information Geometry*, 1(2):137–179, 2018. ISSN 2511-249X. doi: 10.1007/s41884-018-0014-4. URL https://doi.org/10.1007/s41884-018-0014-4.

[73] Ivan Marisca, Jacob Bamberger, Cesare Alippi, and Michael M. Bronstein. Over-squashing in spatiotemporal graph neural networks. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=CVp0WCw4a1.

[74] Peter Mostowsky, Vincent Dutordoir, Iskander Azangulov, Noémie Jaquier, Michael John Hutchinson, Aditya Ravuri, Leonel Rozo, Alexander Terenin, and Viacheslav Borovitskiy. The GeometricKernels package: Heat and Matérn kernels for geometric learning on manifolds, meshes, and graphs. *Journal of Machine Learning Research*, 26(276):1–14, 2025.

[75] Christoph Müller and Christoph Wockel. Equivalences of smooth and continuous principal bundles with infinite-dimensional structure group. *Advances in Geometry*, 9(4):605–626, 2009. ISSN 1615-715X. doi: 10.1515/advgeom.2009.032. URL http://dx.doi.org/10.1515/ADVGEOM.2009.032.

[76] Liviu I. Nicolaescu. *Lectures on the Geometry of Manifolds*. World Scientific, Singapore, second edition, 2007. ISBN 9789812708533.

[77] Mathilde Papillon, Guillermo Bernardez, Claudio Battiloro, and Nina Miolane. TopoTune: A framework for generalized combinatorial complex neural networks. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 47924–47952, 2025.

[78] Yuhan Peng, Junwen Dong, Yuzhi Zeng, Hao Li, Ce Ju, Huitao Feng, Diaaeldin Taha, Anna Wienhard, and Kelin Xia. Sheaf neural networks on SPD manifolds: Second-order geometric representation learning. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, pp. 97981–98013, 2026.

[79] Peter Petersen. *Riemannian Geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2006. ISBN 978-0-387-29403-2. doi: 10.1007/978-0-387-29403-2.

[80] I. F. Pinelis. Inequalities for distributions of sums of independent random vectors and their application to estimating a density. *Theory of Probability & Its Applications*, 35(3):605–607, 1991. doi: 10.1137/1135088.

[81] Javier Porras-Valenzuela, Zhiyang Wang, Teresa Shang, Yusu Wang, and Alejandro Ribeiro. Size transferability of graph transformers with convolutional positional encodings, 2026. URL https://arxiv.org/abs/2602.15239.

[82] Michael Reed and Barry Simon. *Functional analysis*, volume 1 of *Methods of Modern Mathematical Physics*. Academic Press, New York, 1972. ISBN 0125850018.

[83] Hans Riess and Robert Ghrist. Diffusion of information on networked lattices by gossip. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 5946–5952. IEEE, 2022.

[84] Hans Matthew Riess and Jakob Hansen. Multidimensional persistence module classification via lattice-theoretic convolutions. In *NeurIPS Workshop: TDA & Beyond*, 2020.

[85] Raif M Rustamov et al. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on geometry processing*, volume 257, pp. 225–233, 2007.

[86] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[87] Stefan C Schonsheck, Bin Dong, and Rongjie Lai. Parallel transport convolution: Deformable convolutional networks on manifold-structured data. *SIAM Journal on Imaging Sciences*, 15 (1):367–386, 2022.

[88] Jean-Pierre Serre. Faisceaux algébriques cohérents. *Annals of Mathematics*, 61(2):197–278, 1955.

[89] Lingxuan Shao, Zhenhua Lin, and Fang Yao. Intrinsic riemannian functional data analysis for sparse longitudinal observations. *The Annals of Statistics*, 50(3):1696–1721, 2022.

[90] Allen Dudley Shepard. *A cellular description of the derived category of a stratified space*. Brown University, 1985.

[91] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

[92] Amit Singer and H-T Wu. Vector diffusion maps and the connection laplacian. *Communications on pure and applied mathematics*, 65(8):1067–1144, 2012.

[93] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.

[94] Federica Spoto, Alessia Caponera, and Pierpaolo Brutti. Change point detection for functional autoregressive processes on the sphere. *arXiv preprint arXiv:2512.03255*, 2025.

[95] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

[96] Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Mehdi Azabou, Eva L Dyer, Remi Munos, Petar Veličković, and Michal Valko. Large-scale representation learning on graphs via bootstrapping. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=0UXT6PpRpW.

[97] B. Vallet and B. Lévy. Spectral geometry processing with manifold harmonics. *Computer Graphics Forum*, 27(2):251–260, 2008. ISSN 1467-8659. doi: 10.1111/j.1467-8659.2008.01122.x.

[98] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[99] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[100] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pp. 555–586, 2008.

[101] Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Stability of neural networks on riemannian manifolds. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 1845–1849. IEEE, 2021.

[102] Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Convolutional neural networks on manifolds: From graphs and back. In *2022 56th Asilomar Conference on Signals, Systems, and Computers*, pp. 356–360. IEEE, 2022.

[103] Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Geometric graph filters and neural networks: Limit properties and discriminability trade-offs. *IEEE Transactions on Signal Processing*, 72:2244–2259, 2024.

[104] Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Stability to deformations of manifold filters and manifold neural networks. *IEEE Transactions on Signal Processing*, 72:2130–2146, 2024.

[105] Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. *Equivariant and Coordinate Independent Convolutional Networks: A Gauge Field Theory of Neural Networks*. World Scientific, 2026.

[106] Hao Wu, Kangyu Weng, Shuyi Zhou, Xiaomeng Huang, and Wei Xiong. Neural manifold operators for learning the evolution of physical dynamics. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3356–3366, 2024.

[107] Yuxing Xie, Jiaojiao Tian, and Xiao Xiang Zhu. Linking points with labels in 3d: A review of point cloud semantic segmentation. *IEEE Geoscience and remote sensing magazine*, 8(4): 38–59, 2020.

[108] Zhenyu Yang, Shuo Huang, Han Feng, and Ding-Xuan Zhou. Spherical analysis of learning nonlinear functionals. *Constructive Approximation*, pp. 1–29, 2026.

[109] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

[110] Jaemyung Yu, Jaehyun Choi, Dong-Jae Lee, HyeongGwon Hong, and Junmo Kim. Self-supervised transformation learning for equivariant representations. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=87AXdbkRyd`.

[111] Olga Zaghen, Floor Eijkelboom, Alison Pouplin, and Erik J Bekkers. Towards variational flow matching on general geometries. In *ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy*, 2025.

[112] Yanan Zhao, Feng Ji, Xingchao Jian, and Wee Peng Tay. Graph distribution-valued signals: A wasserstein space perspective. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 166–170. IEEE, 2026.

Appendix Contents

A Extended Related Works 18

B Broader Impact and Future Directions 19

C Existing Convolutional Architectures as Actualizations of HilbNets 20

C.1 Universality of HilbNets 20
C.1.1 CNNs, GNNs, and Sheaf NNs 20
C.1.2 Equivariant CNNs and GNNs 21
C.1.3 Spatio-Temporal GNNs 21

D Practical Implementations of Parallel Transport 21

E Parallel Transport Parametrizations 22

E.1 From Bundle Transports to Network-Sheaf Restrictions 22
E.2 Transport Hypothesis Classes 23
E.3 Parameter Counts 25

F Additional Experimental Details 26

F.1 Synthetic Experiments: The Statistical Bundle over Centered Gaussians 26
F.1.1 The Bundle 26
F.1.2 Levi-Civita Connection and Closed-Form Parallel Transport 26
F.1.3 Cholesky Rescaling 27
F.1.4 Sample Construction 27
F.1.5 Spectral Stability under Sampling Density Increase 28
F.1.6 Hyperparameters 30
F.2 Traffic Forecasting: Experimental Details 30
F.2.1 Datasets 30
F.2.2 Task Formulation 30
F.2.3 Model Variants and Baselines 31
F.2.4 Detailed Results 31
F.2.5 Hyperparameters 31

G Mathematical Background 32

G.1 Hilbert Bundles 32
G.2 Connection Laplacian .....
G.3 Heat Flow on a Hilbert Bundle 37
G.4 Borel Functional Calculus 38
G.5 Cellular Sheaves and Sheaf Laplacians 39
G.6 Empirical Laplacians 41

H Proofs of Results 42

H.1 Auxiliary Lemmas for Theorem 1 .....
H.2 Key Lemmas for Theorem 1 .....
H.3 Proof of Theorem 1 47
H.4 Key Lemmas for Theorem 2 .....
H.5 Proof of Theorem 2 49
H.6 Key Lemmas for Corollary 1 (Convergence in Architecture) .....
H.7 Proof of Corollary 1 (Convergence in Architecture) .....
H.8 Proof of Corollary 2 (Transferability) .....

A Extended Related Works

The connection between (possibly) continuous domains (manifolds and bundles) and discrete structure (graphs and cellular sheaves) first emerged in pioneering investigations on the so-called manifold hypothesis. This hypothesis posits that, although data may live in a high-dimensional ambient space, they are effectively generated by sampling from one or several low-dimensional (Riemannian) manifolds [21]. The manifold hypothesis underpins several modern spectral graph methods, i.e., nonlinear dimensionality-reduction/clustering/(deep) learning techniques that exploit latent geometric structures. The renowned work [14] from Belkin & Niyogi proved that, assuming access to a finite

point cloud (the signals) sampled from the underlying manifold, it is possible to build a weighted undirected graph whose Laplacian converges to the Laplace-Beltrami operator of the underlying manifold in probability as the number of samples goes to infinity.

The work in [14] and related results, such as [92, 93], have been used (directly or indirectly) to design principled learning systems over manifolds. A consistent fraction of this literature focused on scalar manifold signals, thus the case in which one or more scalar values are attached to each point of a manifold. Notable examples are manifold convolutional neural networks [102, 67], kernel methods and Gaussian processes on manifolds [20, 74], as well as a growing literature of generative manifold models [111, 33]. In a complementary direction, operator-learning methods on manifolds extend neural operators beyond Euclidean domains; these methods handle an infinite-dimensional object globally, but they still assign a finite vector to each point of a manifold. Most of the works in this class are instances of neural manifold operators [26, 19, 106, 58], which aim at resolution-independent learning of PDE solution operators. Some works explored vector-valued manifold signals, i.e., multivariate real-valued functions supported on manifolds; in this case, one or more finite vectors are attached to each point of a manifold. Examples are tangent bundle convolutional neural networks [11] and vector-field Gaussian processes on manifolds built via gauge-independent projected kernels [56]. Moreover, especially in the statistics community, functional observations with manifold structure, i.e., manifold-valued functions supported on the real line or real-valued functions supported on manifolds, have been long studied [70, 31, 89], and recent works have started analyzing autoregressive processes on the sphere [24, 94]. Finally, learning systems acting on discrete bundles, i.e., bundles whose base space is a finite set/discrete manifold, have been recently investigated [5]. Despite their diversity, all the models cited in this paragraph use finite-dimensional fibers and implicitly assume the Levi-Civita connection. As such, they do not allow for the arbitrary connections or the potentially infinite-dimensional signals considered. One of the main reasons behind this gap is the lack of a rigorous generalization of the [14]'s convergence result in these settings, which is our main contribution.

Pioneering works on sheaf theory can be found in [65, 88, 48]. *Cellular sheaves* are combinatorial instances of sheaves that have been introduced in [90] and later rediscovered in [29]. In [90, 29], these sheaves were first defined over regular cell complexes, hence the term "cellular" sheaves. However, as in this work, cellular sheaves are often defined over tamer objects, here graphs. In [52, 35], the authors studied the problem of learning *vector cellular sheaves*, i.e., cellular sheaves over undirected graphs with finite-dimensional node signals. The works in [53, 51, 42, 83] introduced a novel class of diffusion dynamics on vector cellular sheaves. In [18, 50, 7, 39, 37, 78, 9, 11], neural networks operating on vector cellular sheaves over (undirected, directed, hyper) graphs with finite-dimensional signals are presented, generalizing graph neural networks. We again note however, that all these works implicitly or explicitly restrict to consider either the Levi-Civita or flat connections. Additionally, the work in [18] exploited vector cellular sheaf theory to show that the underlying geometry of the graph gives rise to oversmoothing behavior of GCNs. Also, (vector and general) cellular sheaves recently appeared in causal theory [30], control [49], and telecommunications [47]. Finally, the works in [9, 11] showed that neural networks for tangent bundle signals can be implemented as certain sheaf neural networks operating on vector cellular sheaves from manifold samples.

B Broader Impact and Future Directions

The potential impact of this work extends well beyond the effectiveness of the HilbNet architecture. Our convergence result unifies and extends the graph- and vector-diffusion convergence theories of [14, 93], thereby enabling novel geometric learning systems for genuinely infinite-dimensional manifold-supported and equipped with arbitrary connections. HilbNets are just a first (principled, transferable) instance of such systems, but our result opens several new avenues.

Clustering and Dimensionality Reduction. Classical Laplacian-based methods for clustering [100] and nonlinear dimensionality reduction [13, 28] all rely, either explicitly or implicitly, on the convergence of the graph Laplacian to the Laplace–Beltrami operator. Our Theorem 1 provides the analogous foundation in the Hilbert bundle setting, immediately suggesting sheaf-spectral generalizations of these techniques. For instance, one may define Hilbert sheaf eigenmaps by computing the leading eigensections of $\Delta_{\mathcal{F}_n^t}$ and using them as coordinates, yielding embeddings that are aware not only of the base manifold geometry but also of the fiber-wise coupling encoded by the connection. This is particularly promising for data such as spatiotemporal fields or distributional signals, where standard spectral methods discard the internal structure of each observation. Similarly, sheaf-spectral clustering would partition data by jointly considering geometric proximity on $\mathcal{M}$ and coherence of the infinite-dimensional signals across fibers, a strictly richer criterion than what scalar graph

Laplacians can capture. The finite-rank convergence guarantee of Theorem 2 ensures that such methods can be implemented with truncated signals while remaining provably consistent with the underlying continuous geometry.

Structured Self-Supervised Learning. Self-supervised learning (SSL) has largely been built around objectives that encourage invariance or equivariance with respect to augmentations of the base domain [109, 96, 32, 110]. Our framework suggests a more structured family of SSL methods. Because the Hilbert sheaf Laplacian encodes both spatial geometry and fiber-wise transport, one can design contrastive or non-contrastive objectives that encourage learned representations to be sections of an appropriate bundle, i.e., to satisfy local consistency constraints dictated by the restriction maps. Such objectives would yield representations that are not merely invariant to domain augmentations but are geometrically coherent across fibers, a property that is especially desirable when downstream tasks depend on the relational structure between signals at different manifold points, as in multi-sensor forecasting or multi-agent coordination.

Generalizability Theory and Mechanistic Interpretability for Transformers. Another promising direction concerns the connection between our framework and transformer architectures [98]. In a standard transformer, each token is equipped with a positional encoding, either fixed (e.g., sinusoidal) or learned, that situates it in a continuous geometric space [95]. These positional encodings can be viewed as sampled points on a base manifold $\mathcal{M}$, with the encoding scheme implicitly defining the metric structure of the domain. The residual-stream representation at each position, or, in the infinite-width or infinite-context limit, the full distribution over possible activations, then lives in a Hilbert space fibered over this base point, so that the collection of representations across positions constitutes a section of a Hilbert bundle over $\mathcal{M}$. The attention mechanism then defines a data-dependent transport between these fibers. In this view, a self-attention layer is an instance of a single-step diffusion under a learned sheaf Laplacian whose base graph is determined by the sampled positional encodings. This may in some sense be viewed as a more precise incarnation of the recently introduced *geodesic hypothesis* [55], where the autoregressive output of transformers is modeled by a stochastic diffusion PDE in Euclidean space, rather than the proposed manifold-theoretic treatment. Making this correspondence precise would allow one to import the convergence and transferability machinery of Theorems 1–2 into the transformer setting. We note that there exists a recent line of work that attempts to adapt Laplacian-based GNN generalization and stability results to transformers [81], but due to their fundamental reliance on the convegrence result of [14], must work in a somewhat simplified setting. In conjunction with our convergence result, the generality of the Hilbert sheaf Laplacian is potentially well-suited for establishing extensions of these generalization theorems for a broader class of transformers. On the interpretability side, decomposing attention into a positional-affinity component and a fiber-transport component offers a principled lens through which to study what information each head moves and how it is transformed in transit. One could, for example, measure the holonomy of the learned connection around closed loops of attention to detect whether a head implements a nontrivial geometric transformation. While formalizing these connections requires treatment of the data-dependence of the connection and the interplay between positional and content-based attention, the mathematical infrastructure developed in this work provides a strong starting point.

C Existing Convolutional Architectures as Actualizations of HilbNets

C.1 Universality of HilbNets

Due to the first-principles approach to the construction of HilbNets, we note that they serve as a sort of universal architecture. That is, several popular variations of convolutional architectures in geometric deep learning, across domains and modalities, can be derived as particular instantiations of HilbNets — even when their construction does not explicitly invoke cellular sheaves. We consider a few concrete examples of this philosophy.

C.1.1 CNNs, GNNs, and Sheaf NNs

Convolutional neural networks (CNNs) and graph neural networks (GNNs) are both often formalized as acting on signals $f : \mathcal{M} \to \mathbb{R}$. In particular, by the consistency of the discrete Fourier transform, CNNs operating on fixed grid can be viewed as operating on a principled discretization of $\mathcal{M} = [0, 1]^2$. We may more generally view the uniform grid on which CNNs act as a particular instantiation of a graph, and thus view CNNs as a special case of GNNs [59], and the relevant convolutional operator as the graph Laplacian. The GNNs can likewise be viewed as principled discretizations of manifold

neural networks [104], precisely by the convergence result of [14]. Sheaf neural networks [50, 18] may then be understood as an enrichment that allows for matrix-valued edge weights rather than scalars during convolution. Further, in [9], it was made precise that sheaf neural networks may in particular, be viewed as acting upon tangent bundle signals $s : \mathcal{M} \to T\mathcal{M}$, via the convergence result of [92]. In particular, the convergence result of [92] applies strictly to tangent bundle setting, providing an explanation as to why existing works of sheaf neural networks either explicitly or implicitly restrict to discretizing the tangent bundle with either flat or Levi-Civita connections [7, 5]. As such, existing sheaf neural network architectures can be typically be recovered as HilbNets under the paradigm that $\mathcal{E} = T\mathcal{M}$.

C.1.2 Equivariant CNNs and GNNs

The equivariant case of CNNs and GNNs arises when we wish for our architecture to respect some underlying symmetry group $G \curvearrowright \mathcal{M}$. More generally, there may not exist a global representation of the symmetry, but rather only a local representation. In physics, this is known as a *gauge symmetry*, and is formalized by considering our signal f as a section of a bundle with connection $(\mathcal{M}, \mathcal{E}, \nabla)$, where the group action is then encoded as a symmetry of the connection ∇. As such, *gauge-equivariant* CNNs and GNNs constitute perhaps the most general equivariant architectures in the literature (see Weiler et al. [105] for a thorough introduction in the CNN case), and are formulated precisely as convolutions at the level of sections of a frame bundle (although the necessary datum of a connection is often suppressed in the literature). From this perspective, it is natural that gauge-equivariant CNNs and GNNs can be derived as particular cases of cellular sheaf networks (see the Li et al. [68] for a more in-depth exploration of this perspective). Thus, by noting that these architectures can be equivalently reformulated in the language of sheaves, our Theorem 1, as well as the consequent transferability result, can be seen to apply to these architectures. In particular, this may be understood as establishing the theoretical bedrock for the intuitive idea that as the underlying mesh or graph is increasingly refined, these architectures indeed approach continuous operators on sections of the underlying bundle, while maintaining equivariance across scales.

C.1.3 Spatio-Temporal GNNs

A formal treatment of signal processing of graphs whose signals at each node are timeseries is still emerging and is an active area of research, and these filtering techniques then serve as the basis for the development of spatiotemporal graph neural networks (STGNNs). In this literature, it is common to consider convolutional operators built via the *joint Laplacian* $L_J = L_T \otimes \mathrm{Id}_G + \mathrm{Id}_T \otimes L_G$, where L_G is the graph-domain Laplacian and L_T is the 'time-domain' Laplacian [45]. Due to this decomposition, the resulting 'time' and 'space' filters commute, allowing for the development of both *time-and-space* or *time-then-space* STGNNs [73]. Consider now the continuous setting. Convolutions via Δ_∇ intertwine the spatial and temporal domains, and this is precisely encoded by our parallel transport maps. So suppose our bundle is trivial $\mathcal{E} = \mathcal{M} \times L^2(\mathbb{R}^n)$ with trivial connection ∇. In this case, our parallel transport maps are simply $P_\gamma = Id$, and the connection Laplacian collapses to the Laplace-Beltrami operator. On product manifolds $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, the Laplace-Beltrami operator decomposes as $\Delta_\mathcal{M} = \Delta_{\mathcal{M}_1} \otimes \mathrm{Id}_{\mathcal{M}_2} + \mathrm{Id}_{\mathcal{M}_1} \otimes \Delta_{\mathcal{M}_2}$, implying that heat flow is given by $e^{t\Delta_M} = e^{t\Delta_{M_1}} \otimes e^{t\Delta_{M_2}}$ (see [46] for a formal derivation). Expressing $e^{t\Delta_\mathcal{M}}$ as an integral operator via the heat kernel, the fact that spatial and temporal filters commute in this case is then simply an application of Fubini's theorem. As such, we see that the type of filtering commonly considered in STGNNs is recovered precisely as the 'base case' of HilbNet, and in particular, our robustness guarantees can also be applied to these STGNN architectures.

D Practical Implementations of Parallel Transport

As we have established, a key strength of the HilbNets architecture is the ability to encode signal-level geometric priors through the principled incorporation of relevant parallel transport operators. This naturally raises the question as to how these transport operators should be implemented in-practice. We may consider three general classes of use-cases.

Task-Inherent Priors The most theoretically well-grounded case is when knowledge of the geometry of task itself may be utilized to build our parallel transport operators. For instance, suppose the nodes of our base graph represent cameras and the task is multi-view 3D recognition. Then the relevant transport operators should record the rotation $P_{x_i \to x_j} \in SO(3)$ that aligns views as in [68], resulting in an appropriately equivariant sheaf Laplacian operator. More generally, whenever the data modality lacks a 'global' reference frame or coordinate system, then the appropriate alignments between

local reference frames precisely gives rise to a connection and the associated parallel transport. For instance, biomedical timeseries analysis often utilizes algorithms based upon the large deformation diffeomorphic metric mapping (LDDMM) [12], a core part of which may be understood as extracting the necessary parallel transport from the data using the first-order ODE definition. We may also consider the tangent bundle networks of [11] in this category, as they use explicit vector-field data from which they may then compute the necessary sheaf transition maps via local PCA. As such, we see that HilbNets may be applied to any of these settings, where the relevant parallel transport would be completely determined by the task itself and thus, can typically be explicitly pre-computed.

Domain-Inherent Priors Alternatively, it is often the case that we may not have access to task-specific priors, but rather to general knowledge of the structure of the signal-domain. For instance, in many domains, our generic stalks may be equipped with the additional structure of a reproducing kernel Hilbert space (RKHS), i.e. $\mathcal{H}_\kappa$. Analogously to the previous case, we may then view parallel transport as operators that that maximize alignment, but now with respect to our kernel. For instance, given a choice of similarity kernel κ between timeseries or distributions, then given our initial section data $\{S_u\}_{u=1}^{F_0}$, we may define our parallel transport operator matrices via

$$P_{x_j \to x_i}^{(d,e_{ij})} := \arg\max_{\mathbf{T}\in\mathcal{C}} \sum_{q=1}^{F_0} \kappa(\mathbf{T}\mathbf{S}_{i,q}, \mathbf{S}_{j,q}) \tag{15}$$

for some suitable class of operators $\mathcal{C} \subseteq O(d)$, and force the diagonal blocks of the sheaf Laplacian to be the sum of the scalar edge weights given by the kernel. This is exactly the discretized Sheaf Laplacian from (8). In practice, given a choice of similarity kernel on our fibers, we may then either precompute these parallel transport operators using the above optimization objective or learn them end-to-end with the model's learned filters. In the latter case, (15) is applied as a regularization to the task loss. The special case in which (15) is not employed at all, $\mathcal{C} = O(d)$, $K = 1$ in (10), recovers the sheaf diffusion neural network from [18]. As such, we see that the greater generality of the Hilbert sheaf Laplacian consequently lends itself to more flexible and perhaps more broadly applicable design choices than existing sheaf neural networks.

E Parallel Transport Parametrizations

This appendix details the finite-dimensional transport parametrizations used to instantiate the network sheaf Laplacian $\Delta_{\mathcal{F}_{n,d}^t}$ in the experiments. In particular, this particular instantiation of HilbNets may be considered as a paticular of the end-to-end learning paradigm introduced in D for polynomial filters and a few demonstrative classes of $\mathcal{C}$ and κ. The discussion should be read as a continuation of the two-stage discretization in Section 5: after sampling the manifold, we obtain a Hilbert cellular sheaf $\mathcal{F}_n^t$; after sampling or projecting the fibers, we obtain a finite-dimensional network sheaf $\mathcal{F}_{n,d}^t$ with d-dimensional stalks.

E.1 From bundle transports to network-sheaf restrictions

Recall that, before signal discretization, the Hilbert cellular sheaf $\mathcal{F}_n^t$ induced by a sample $\mathcal{X}_n = \{x_1, \dots, x_n\}$ assigns the node stalk

$$\mathcal{F}_n^t(x_i) = \mathcal{E}_{x_i} \tag{16}$$

and, for an edge $e_{ij} \in$, the edge stalk

$$\mathcal{F}_n^t(e_{ij}) = \mathcal{E}_{m_{\gamma_{ij}}}, \tag{17}$$

where $m_{\gamma_{ij}}$ is the midpoint of the chosen geodesic between x_i and x_j. Its restriction maps are weighted parallel transports of the form

$$(\mathcal{F}_n^t)_{x_i \le e_{ij}} = \sqrt{k_{ij}^t}\, P_{x_i \to m_{\gamma_{ij}}}, \qquad k_{ij}^t = \exp\left(-\frac{d_{\mathcal{M}}(x_i, x_j)^2}{4t}\right). \tag{18}$$

After fiber discretization, the network sheaf $\mathcal{F}_{n,d}^t$ has finite-dimensional stalks, which we identify with $\mathbb{R}^d$ after choosing the first d basis elements of the fiber Hilbert space. The corresponding restriction maps are matrices

$$(\mathcal{F}_{n,d}^t)_{x_i \le e_{ij}} : \mathbb{R}^d \to \mathbb{R}^d. \tag{19}$$

For the pragmatic parametrizations used in the experiments, it is useful to express the same sheaf Laplacian in node-to-node transport coordinates. Fix an orientation convention for each edge e_{ij}.

After identifying the edge stalk with the coordinate system of one endpoint, we write the restrictions as

$$(\mathcal{F}^t_{n,d})_{x_i \le e_{ij}} = \sqrt{k^t_{ij}}\, I_d, \qquad (\mathcal{F}^t_{n,d})_{x_j \le e_{ij}} = \sqrt{k^t_{ij}}\, P^{(d,e_{ij})}_{x_j \to x_i}, \tag{20}$$

where

$$P^{(d,e_{ij})}_{x_j \to x_i} \in O(d) \tag{21}$$

is the finite-dimensional transport carrying the discretized fiber over x_j into the discretized fiber over x_i along edge e_{ij}. When the transport comes from the continuous Hilbert bundle, this matrix is the finite-dimensional representation of the corresponding parallel transport, after the chosen fiber projection and coordinate identification. When the continuous connection is unknown, $P^{(d,e_{ij})}_{x_j \to x_i}$ is instead chosen from a transport hypothesis class.

With the shorthand

$$P^e_{j \to i} := P^{(d,e_{ij})}_{x_j \to x_i}, \tag{22}$$

the action of the network sheaf Laplacian on a sampled signal

$$\mathbf{s}_{n,d} = (\mathbf{s}_{x_1}, \dots, \mathbf{s}_{x_n}) \in C^0(\mathcal{F}^t_{n,d}; G_n), \qquad \mathbf{s}_{x_i} \in \mathbb{R}^d, \tag{23}$$

takes the concrete form

$$(\Delta_{\mathcal{F}^t_{n,d}} \mathbf{s}_{n,d})_{x_i} = \sum_{x_j \in \mathcal{N}(x_i)} k^t_{ij} \left(\mathbf{s}_{x_i} - P^{e_{ij}}_{j \to i} \mathbf{s}_{x_j} \right). \tag{24}$$

Equivalently, $\Delta_{\mathcal{F}^t_{n,d}} \in \mathbb{R}^{nd \times nd}$ is the block matrix with blocks

$$(\Delta_{\mathcal{F}^t_{n,d}})_{ij} = \begin{cases} \sum_{r:\, e_{ir} \in E} k^t_{ir} I_d, & i = j, \\ -k^t_{ij} P^{e_{ij}}_{j \to i}, & i \neq j \text{ and } e_{ij} \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

This is the same sheaf Laplacian defined in Section 4, specialized to the coordinate convention in (20). The scalar weight k^t_{ij} controls how strongly the two sampled base points interact, while the transport matrix $P^{e_{ij}}_{j \to i}$ controls how the two discretized fibers are aligned before their signals are compared.

E.2 Transport hypothesis classes

As mentioned in D, the true connection is often unknown and parallel transport maps must be learned. In the experimental results of this work, we always learn the parallel transport maps end-to-end using the task loss regularized with (15), and we restrict each edgewise transport to hypothesis classes $\mathcal{C}$ such that

$$P^{e_{ij}}_{j \to i} \in \mathcal{C} \subseteq O(d). \tag{26}$$

We use three transport classes in the experiments: frozen identity, free orthogonal transports, and circulant or time-stationary transports.

Frozen identity. The simplest class is

$$\mathcal{C}_{\mathrm{id}} = \{I_d\}. \tag{27}$$

This recovers the usual assumption that neighboring fibers are canonically identified and that no non-trivial alignment is needed. In this case,

$$P^{e_{ij}}_{j \to i} = I_d \tag{28}$$

for every edge, and (24) becomes

$$(\Delta_{\mathcal{F}^t_{n,d}} \mathbf{s}_{n,d})_{x_i} = \sum_{x_j \in \mathcal{N}(x_i)} k^t_{ij} \left(\mathbf{s}_{x_i} - \mathbf{s}_{x_j} \right). \tag{29}$$

Thus, frozen identity reduces the sheaf Laplacian to a standard weighted graph Laplacian applied independently to each fiber coordinate, and, therefore, HilbNets to standard GCNs. It is a useful baseline: any improvement over frozen identity quantifies the value of learning or imposing non-trivial transports.

Free orthogonal transports. The most expressive finite-dimensional class is

$$\mathcal{C}_{\text{free}} = O(d), \tag{30}$$

or, when the target transports are known to lie in the identity component,

$$\mathcal{C}_{\text{free}} = SO(d). \tag{31}$$

In the experiments, we parameterize free orthogonal transports by products of Householder reflections. For a nonzero vector $v \in \mathbb{R}^d$, define the Householder reflection

$$H(v) = I_d - 2\frac{vv^\top}{\|v\|_2^2}. \tag{32}$$

Each $H(v)$ is orthogonal and symmetric:

$$H(v)^\top H(v) = I_d, \qquad H(v)^\top = H(v). \tag{33}$$

For each oriented edge e_{ij}, we store R Householder vectors

$$v_{e_{ij},1}, \dots, v_{e_{ij},R} \in \mathbb{R}^d \tag{34}$$

and define

$$P_{j\to i}^{e_{ij}} = H(v_{e_{ij},R}) \cdots H(v_{e_{ij},1}). \tag{35}$$

Therefore $P_{j\to i}^{e_{ij}}$ is exactly orthogonal for every parameter value.

By the Cartan–Dieudonné theorem, every matrix in $O(d)$ can be represented as a product of at most d Householder reflections. In practice, the choice of R is dataset-dependent: in our synthetic experiments we use $R = 16$ for $d = m = 10$, a modest over-parametrization that aids optimization in our traffic experiments we use $R = 8$ for $d = T = 12$, which parameterizes a strict Householder subset of $O(T)$ rather than the full orthogonal group, and which we find sufficient for the alignment patterns observed in the data. If one fixes exactly R non-degenerate reflections, the determinant parity is fixed:

$$\det(P_{j\to i}^{e_{ij}}) = (-1)^R. \tag{36}$$

Thus, an even number of reflections parameterizes the identity component $SO(d)$, while an odd number parameterizes the other component. In the synthetic Gaussian experiment, the ground-truth Levi-Civita transports are obtained continuously from the identity along geodesics and, after Cholesky rescaling, lie in the identity component. Hence an even number of reflections is appropriate. If both connected components of $O(d)$ are needed, one may add a fixed final reflection or a discrete sign component. For numerical stability, the implementation uses

$$H(v_{e_{ij},r}) = I_d - 2\frac{v_{e_{ij},r} v_{e_{ij},r}^\top}{\|v_{e_{ij},r}\|_2^2 + \epsilon}, \tag{37}$$

with a small $\epsilon > 0$ to avoid division by zero at degenerate $v_{e_{ij},r}$. This recovers the exact Householder reflection $H(q) = I_d - 2qq^\top$ with $q = v_{e_{ij},r}/\|v_{e_{ij},r}\|_2$ in the limit $\epsilon \to 0$ for $\|v_{e_{ij},r}\|_2 > 0$, and yields a matrix orthogonal up to numerical precision.

The free class is useful as an expressivity test. If the target transport belongs to $O(d)$ in the chosen coordinates, then the Householder class can represent it. This is precisely the role it plays in the synthetic statistical-bundle experiment, where Cholesky rescaling converts the intrinsic Wasserstein-unitary Levi-Civita transports into Euclidean-orthogonal matrices. In real-data experiments, the free class serves as a high-capacity transport baseline.

Circulant or time-stationary transports. For time-series fibers, the discretized fiber dimension is the number of retained time samples, so we write $d = T$. A natural prior is that inter-fiber transport should commute with time shifts. Let

$$\mathsf{S}_T : \mathbb{R}^T \to \mathbb{R}^T \tag{38}$$

be the cyclic shift operator. A time-stationary transport is one satisfying

$$P_{j\to i}^{e_{ij}} \mathsf{S}_T = \mathsf{S}_T P_{j\to i}^{e_{ij}}. \tag{39}$$

The commutant of the cyclic shift is the algebra of circulant matrices. Requiring in addition that $P_{j\to i}^{e_{ij}}$ be orthogonal gives the class of orthogonal circulant transports.

Let F_T denote the unitary discrete Fourier transform matrix. Then every orthogonal circulant transport has the form

$$P_{j\to i}^{e_{ij}} = F_T^* \operatorname{diag}(\lambda_{e_{ij}}) F_T, \qquad |\lambda_{e_{ij},k}| = 1. \tag{40}$$

For real-valued time-domain signals, the Fourier multipliers must satisfy conjugate symmetry:

$$\lambda_{e_{ij},T-k} = \overline{\lambda_{e_{ij},k}}. \tag{41}$$

We therefore store only the independent positive-frequency phases. Let

$$m_T = \left\lfloor \frac{T-1}{2} \right\rfloor. \tag{42}$$

The learnable parameter for edge e_{ij} is

$$\varphi_{e_{ij}} = (\varphi_{e_{ij},1}, \ldots, \varphi_{e_{ij},m_T}) \in \mathbb{R}^{m_T}. \tag{43}$$

We define

$$\lambda_{e_{ij},0} = 1, \qquad \lambda_{e_{ij},k} = e^{i\varphi_{e_{ij},k}}, \qquad \lambda_{e_{ij},T-k} = e^{-i\varphi_{e_{ij},k}}, \quad k = 1, \ldots, m_T. \tag{44}$$

If T is even, the Nyquist frequency is self-conjugate and is fixed to

$$\lambda_{e_{ij},T/2} = 1 \tag{45}$$

for the identity-component parametrization. This yields a real orthogonal circulant matrix through (40). Equivalently, $P_{j\to i}^{e_{ij}}$ can be constructed in real arithmetic from its first column. For $r = 0, \ldots, T-1$, define

$$c_{e_{ij}}[r] = \frac{1}{T}\left[1 + \mathbb{I}_{\{T \text{ even}\}}(-1)^r + 2\sum_{k=1}^{m_T} \cos\left(\varphi_{e_{ij},k} + \frac{2\pi k r}{T}\right)\right]. \tag{46}$$

The full circulant matrix is then

$$(P_{j\to i}^{e_{ij}})_{ab} = c_{e_{ij}}[(a-b) \bmod T], \qquad a, b = 0, \ldots, T-1. \tag{47}$$

This form is convenient for implementation because it avoids explicitly manipulating complex-valued matrices. The circulant class has only

$$m_T = \left\lfloor \frac{T-1}{2} \right\rfloor \tag{48}$$

parameters per edge, compared with $d(d-1)/2$ degrees of freedom for a general orthogonal matrix. Each phase $\varphi_{e_{ij},k}$ has a direct interpretation as the phase lag at frequency k between the two endpoint fibers. Thus, the transport may advance or delay oscillatory components across an edge, but it cannot arbitrarily mix frequencies or reshape the waveform. This is the intended inductive bias for spatiotemporal signals such as traffic or sensor time series, where neighboring sensors may observe delayed or phase-shifted versions of related temporal patterns. In the synthetic experiment, the same class is used more abstractly as a structured subgroup of $O(d)$ against which the ground-truth transports can be projected.

E.3 Parameter counts

For a graph $G_n = (\mathcal{X}_n, E)$ with $|E|$ undirected edges and discretized fiber dimension d, the transport parameter counts are summarized in Table 3.

Transport class	Parameters per edge	Interpretation
Frozen identity	0	no learned alignment
Free Householder	Rd	product of R reflections in $O(d)$
Full circulant	$\lfloor (d-1)/2 \rfloor$	one phase per positive frequency

Table 3: Parameter counts for the finite-dimensional transport classes used in $\mathcal{F}_{n,d}^t$.

Thus, the free class is maximally expressive but parameter-heavy, while the circulant classes encode a strong time-stationary prior and scale linearly with the number of frequencies or bands.

F Additional Experimental Details

F.1 Synthetic experiments: the statistical bundle over centered Gaussians

F.1.1 The bundle

The base manifold is $\mathcal{M} = \mathrm{Sym}^{++}(p)$, the open cone of $p \times p$ symmetric positive-definite matrices. Each $\Sigma \in \mathcal{M}$ parameterizes a centered Gaussian $\mathcal{N}(0, \Sigma)$ on $\mathbb{R}^p$ with density ρ_Σ. We equip $\mathcal{M}$ with the Otto-Wasserstein metric, namely the Riemannian metric induced on $\mathrm{Sym}^{++}(p)$ by the optimal-transport distance W_2 between centered Gaussian measures.

Concretely, the tangent space is naturally identified with symmetric matrices,

$$T_\Sigma \mathcal{M} \cong \mathrm{Sym}(p), \tag{49}$$

and the Otto-Wasserstein inner product between $U, V \in \mathrm{Sym}(p)$ is

$$W_\Sigma(U, V) = \tfrac{1}{2} \operatorname{Tr}\big(L_\Sigma[U]\, V\big), \qquad L_\Sigma[U]\Sigma + \Sigma L_\Sigma[U] = U, \tag{50}$$

where $L_\Sigma[U]$ is the unique symmetric solution of the Lyapunov equation.

Above each Σ, the ambient Hilbert fiber is the vector-field space

$$\mathcal{H}_\Sigma := L^2(\rho_\Sigma; \mathbb{R}^p), \tag{51}$$

equipped with the inner product

$$\langle a, b \rangle_{\mathcal{H}_\Sigma} = \int_{\mathbb{R}^p} a(x)^\top b(x) \rho_\Sigma(x)\, dx. \tag{52}$$

This fiber is genuinely infinite-dimensional. The finite-rank fiber used in the synthetic experiments is the Otto-velocity image of covariance perturbations:

$$\mathcal{E}_\Sigma := \big\{ v_V(x) = L_\Sigma[V]x : V \in \mathrm{Sym}(p) \big\} \subset L^2(\rho_\Sigma; \mathbb{R}^p). \tag{53}$$

Thus, the computational fiber $\mathcal{E}_\Sigma$ is a finite-dimensional statistical subspace of the ambient Hilbert fiber. The map $V \mapsto v_V$ is an isometry between $\mathrm{Sym}(p)$ with the Otto-Wasserstein metric and $\mathcal{E}_\Sigma$ with the $L^2(\rho_\Sigma; \mathbb{R}^p)$ inner product. Indeed, for $U, V \in \mathrm{Sym}(p)$,

$$\langle v_U, v_V \rangle_{\mathcal{H}_\Sigma} = \mathbb{E}_{x \sim \mathcal{N}(0,\Sigma)} \big[x^\top L_\Sigma[U] L_\Sigma[V] x \big] = \operatorname{Tr}\big(L_\Sigma[U] L_\Sigma[V] \Sigma \big). \tag{54}$$

Using $V = L_\Sigma[V]\Sigma + \Sigma L_\Sigma[V]$ and the fact that $L_\Sigma[U]$, $L_\Sigma[V]$, and Σ are symmetric, this equals

$$\tfrac{1}{2} \operatorname{Tr}\big(L_\Sigma[U] V \big) = W_\Sigma(U, V). \tag{55}$$

Therefore,

$$\dim \mathcal{E}_\Sigma = \dim \mathrm{Sym}(p) = d = p(p+1)/2. \tag{56}$$

Since the fibers $\mathcal{E}_\Sigma$ are already m-dimensional, the fiber discretization of Proposition 1 is exact with $d = m$. Throughout this appendix, we therefore write $d = m$ and use the network sheaf notation $\mathcal{F}^t_{n,d}$ and Laplacian $\Delta_{\mathcal{F}^t_{n,d}} \in \mathbb{R}^{nd \times nd}$ from Section 5.

This construction is useful because it gives a faithful but tractable proxy for the Hilbert-bundle settings that motivate HilbNets. It is *faithful* in the sense that the ambient fibers $L^2(\rho_\Sigma; \mathbb{R}^p)$ are infinite-dimensional vector-field Hilbert spaces, and the Levi-Civita connection of $(\mathrm{Sym}^{++}(p), W_\Sigma)$ yields non-trivial, metric-compatible parallel transports. It is *tractable* because the Otto-velocity map selects a finite-rank statistical sub-bundle on which the metric, parallel-transport ODE, and projection of ground-truth transports onto restricted transport classes admit closed-form numerical evaluation. Thus, the experiments probe the finite-rank computational slice used by the implementation.

F.1.2 Levi-Civita connection and closed-form parallel transport

The Levi-Civita connection on $(\mathrm{Sym}^{++}(p), W_\Sigma)$ is the canonical metric-compatible torsion-free connection associated with the Otto-Wasserstein metric. In covariance coordinates, its Christoffel symbol is

$$\Gamma_\Sigma(U, V) = \tfrac{1}{2}\big(U L_\Sigma[V] + V L_\Sigma[U] \big) \in \mathrm{Sym}(p), \tag{57}$$

which is symmetric in (U, V), as required for a torsion-free connection.

Let Σ_t denote the Wasserstein geodesic from Σ_0 to Σ_1. Parallel transport of a tangent vector $V(t) \in \mathrm{Sym}(p)$ along Σ_t is governed by

$$\dot{V}(t) = -\Gamma_{\Sigma_t}\big(\dot{\Sigma}_t, V(t)\big), \qquad V(0) = V_0. \tag{58}$$

We solve this ODE numerically by Euler integration with 50 steps. The resulting linear map on the finite-rank fiber is the ground-truth Levi-Civita transport

$$P^{LC}_{x_i \to x_j} : \mathrm{Sym}(p) \to \mathrm{Sym}(p). \tag{59}$$

In the notation of Def, 3, the restriction maps of the induced sheaf $\mathcal{F}^t_{n,d}$ use the midpoint transports $P^{LC}_{x_i \to m_{ij}}$, which play the role of the discretized parallel transport $P^{(d)}_{x_i \to m_{ij}}$ with $d = m$. These midpoint transports define the restriction maps used in the spectral-stability experiments. The transport-recovery experiments instead regress against the Cholesky-rescaled node-to-node transport $\tilde{P}^{LC}_{x_j \to x_i}$, which adopts the orientation convention of Appendix D.

Because the connection is metric-compatible, $P^{LC}_{x_i \to x_j}$ is unitary with respect to the Otto-Wasserstein inner products on the source and target fibers:

$$W_{x_j}\big(P^{LC}_{x_i \to x_j} U, P^{LC}_{x_i \to x_j} V\big) = W_{x_i}(U, V). \tag{60}$$

We verify this numerically to within 0.5% using 200 Euler steps. This Wasserstein-unitarity is the geometric invariant that justifies comparing the ground-truth transports to orthogonal parametrizations after metric rescaling.

F.1.3 Cholesky rescaling

The free-$O(m)$ transport class used in the implementation is Euclidean-orthogonal in vectorized fiber coordinates (cf. the hypothesis classes in Appendix D with $d = m$). However, the intrinsic fiber metric is W_Σ, not the raw Frobenius metric on $\mathrm{Sym}(p)$. Therefore, $P^{LC}_{x_i \to x_j}$ is not generally orthogonal in raw coordinates.

Let G_Σ be the Gram matrix of W_Σ in a fixed basis of $\mathrm{Sym}(p)$. We factor

$$G_\Sigma = R_\Sigma^\top R_\Sigma \tag{61}$$

by Cholesky decomposition and represent a fiber coordinate vector u in the rescaled frame as

$$\tilde{u} = R_\Sigma u. \tag{62}$$

In this frame, the rescaled Levi-Civita transport is

$$\tilde{P}^{LC}_{x_i \to x_j} = R_{x_j} P^{LC}_{x_i \to x_j} R_{x_i}^{-1}. \tag{63}$$

By Wasserstein-unitarity, it satisfies

$$\big(\tilde{P}^{LC}_{x_i \to x_j}\big)^\top \tilde{P}^{LC}_{x_i \to x_j} = I_m. \tag{64}$$

Hence,

$$\tilde{P}^{LC}_{x_i \to x_j} \in O(m). \tag{65}$$

This is the coordinate system used in the transport-recovery experiments. In these coordinates, the free-$O(m)$ Householder class described in Appendix D contains the ground-truth transports. The spectral-stability metrics are computed from the assembled sheaf Laplacian $\Delta_{\mathcal{F}^t_{n,d}}$ and are invariant to this coordinate choice up to similarity transformation.

F.1.4 Sample construction

We draw samples

$$\Sigma_i = R_i D_i R_i^\top, \qquad i = 1, \ldots, n, \tag{66}$$

where R_i is a Haar-random $p \times p$ orthogonal matrix, obtained by QR decomposition of a standard-normal matrix, and D_i is diagonal with log-uniform spectrum on $[\log 0.5, \log 2.0]$. Equivalently, the eigenvalues of Σ_i lie in $[0.5, 2.0]$ on a log-uniform scale.

Although $\mathrm{Sym}^{++}(p)$ is non-compact, this procedure samples a bounded subset of it. This is appropriate for the finite deployment-regime stability tests reported here, but it is not the normalized-volume sampling assumption used in the asymptotic convergence theorem.

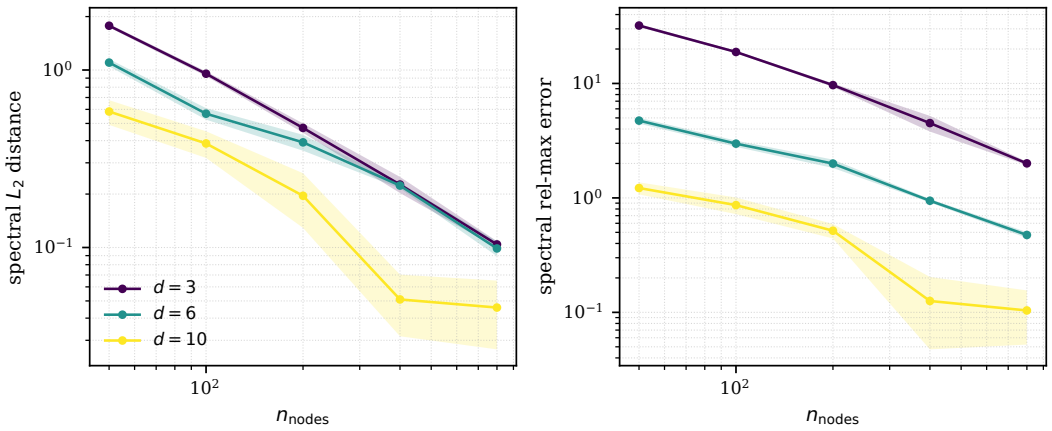

Figure 3: Spectral stability of $\Delta_{\mathcal{F}^t_{n,d}}$ for different signal and manifold sampling densitites. **Left**: aggregate ℓ_2 eigenvalues discrepancy. **Right**: worst-case relative error.

We build a kNN graph $G_n = (\mathcal{X}_n, E)$ with $k = 8$ under the Gaussian Wasserstein distance $W_2(\Sigma_i, \Sigma_j)$ and assign Gaussian-kernel weights

$$k_{ij}^t = \exp\left(-\frac{W_2(\Sigma_i, \Sigma_j)^2}{4t}\right), \qquad t = 0.5. \tag{67}$$

The induced network sheaf $\mathcal{F}^t_{n,d}$ has per-edge restriction maps

$$(\mathcal{F}^t_{n,d})_{x_i \trianglelefteq e_{ij}} = \sqrt{k_{ij}^t}\, P^{LC}_{x_i \to m_{ij}}, \tag{68}$$

as in Def, 3, where m_{ij} is the Wasserstein-geodesic midpoint between Σ_i and Σ_j.

F.1.5 Spectral stability under sampling density increase

For each Gaussian dimension p, sample size $n \in \{50, 100, 200, 400, 800\}$, and random seed, we sample $\mathcal{X}_n \subset \mathrm{Sym}^{++}(p)$, build $\mathcal{F}^t_{n,d}$ using the closed-form Levi-Civita transports, and assemble the sheaf Laplacian

$$\Delta_{\mathcal{F}^t_{n,d}} \in \mathbb{R}^{nd \times nd} \tag{69}$$

as a sparse block matrix. Since the sheaf Laplacian has size $n_{\max} m \times n_{\max} m$ and we require a dense eigendecomposition, we choose $n_{\max}$ so that $n_{\max} \cdot d \approx 10^4$ remains tractable on a single CPU node, yielding $n_{\max} \in \{4000, 2000, 1000\}$ for $p \in \{2, 3, 4\}$ (i.e., $d \in \{3, 6, 10\}$), respectively.
Let

$$\lambda_1^{(n)} \le \cdots \le \lambda_k^{(n)} \tag{70}$$

denote the bottom-k eigenvalues of $\Delta_{\mathcal{F}^t_{n,d}}$, with $k = 32$, and define $\lambda_i^{(n_{\max})}$ analogously for the reference operator $\Delta_{\mathcal{F}^t_{n_{\max},m}}$. We measure both the aggregate ℓ_2 and worst-case relative spectral discrepancy of the bottom-32 eigenvalues of $\Delta_{\mathcal{F}^t_{n,d}}$ against a high-resolution reference, sweeping $p \in \{2, 3, 4\}$ and $n \in \{50, 100, 200, 400, 800\}$, and averaging over 5 sampling realizations. Fig. 3 shows a monotone decrease for both metrics across all dimensions, demonstrating that the sheaf Laplacian stabilizes as manifold sampling density increases, and faster for higher signal sampling densities. Moreover,

Aggregate discrepancy. Fig. 3 (Left) reports the low-frequency spectral ℓ_2 discrepancy

$$\text{spec-L}_2(n) = \frac{1}{k}\left(\sum_{i=1}^{k}\left(\lambda_i^{(n)} - \lambda_i^{(n_{\max})}\right)^2\right)^{1/2}. \tag{71}$$

We average across three seeds and report $\bar{x} \pm s$ as the shaded band.

Worst-case discrepancy. Fig. 3 (Right) reports the relative max error over the bottom-k eigenvalues:

$$\text{spec-rel-max}(n) = \frac{\max_{1\le i\le k}\left|\lambda_i^{(n)} - \lambda_i^{(n_{\max})}\right|}{\lambda_k^{(n_{\max})}}. \tag{72}$$

This complements the aggregate metric by capturing worst-case low-frequency spectral error.

Overall, Fig. 3 shows a monotone decrease for both metrics across all dimensions, demonstrating that the sheaf Laplacian stabilizes as manifold sampling density increases, and faster for higher signal sampling densities.

We train three transport parametrizations from Appendix D, free $O(d)$ (Householder), circulant, and frozen identity, to recover the Levi-Civita transports in Cholesky-rescaled coordinates by minimizing the per-edge transport-MSE loss

$$\mathcal{L} = \frac{1}{|E|}\sum_{e_{ij}\in E}\mathbb{E}_{V\sim\mathcal{N}(0,I_d)}\left\|P_{j\to i}^{e_{ij}}V - \tilde{P}^{LC}_{x_j\to x_i}V\right\|_2^2. \tag{73}$$

Here $P_{j\to i}^{e_{ij}} \in O(d)$ is the rescaled transport produced by the model for edge e_{ij}, $\tilde{P}^{LC}_{x_j\to x_i}$ is the ground-truth Levi-Civita transport from x_j to x_i in Cholesky-rescaled coordinates, and V is a fresh isotropic Gaussian test vector.

By the trace identity

$$\mathbb{E}_{V\sim\mathcal{N}(0,I_d)}\|AV\|_2^2 = \|A\|_F^2, \tag{74}$$

the population loss equals the mean per-edge squared Frobenius distance

$$\mathcal{L}^* = \frac{1}{|E|}\sum_{e_{ij}\in E}\left\|P_{j\to i}^{e_{ij}} - \tilde{P}^{LC}_{x_j\to x_i}\right\|_F^2. \tag{75}$$

Free $O(d)$. The free transport is parameterized by products of Householder reflections, as detailed in Appendix D. Since $\tilde{P}^{LC}_{x_j\to x_i} \in O(d)$ in the Cholesky-rescaled frame, the free-$O(d)$ class contains the target transport and the minimum population loss is zero. Empirically, we observe $\mathcal{L}^*/d \approx 2\cdot 10^{-7}$ at convergence, confirming recovery to numerical precision.

Restricted classes and analytical projections. For each restricted transport hypothesis class $\mathcal{C} \subseteq O(d)$, the population loss minimum is the mean Frobenius distance from each ground-truth transport $\tilde{P}^{LC}_{x_j\to x_i}$ to its best approximation inside $\mathcal{C}$:

$$\mathcal{L}^*_{\mathcal{C}} = \frac{1}{|E|}\sum_{e_{ij}\in E}\min_{T\in\mathcal{C}}\left\|T - \tilde{P}^{LC}_{x_j\to x_i}\right\|_F^2 = \frac{1}{|E|}\sum_{e_{ij}\in E}\left\|\tilde{P}^{LC}_{x_j\to x_i} - \text{proj}_{\mathcal{C}}\big(\tilde{P}^{LC}_{x_j\to x_i}\big)\right\|_F^2. \tag{76}$$

This, divided by d, is the Theory column in Table 1.

Frozen identity. For $\mathcal{C}_{\text{id}} = \{I_d\}$, the projection is trivial: $\text{proj}_{\{I\}}(\tilde{P}^{LC}_{x_j\to x_i}) = I_d$, so

$$\mathcal{L}^*_{\text{frozen}} = \frac{1}{|E|}\sum_{e_{ij}\in E}\left\|\tilde{P}^{LC}_{x_j\to x_i} - I_d\right\|_F^2. \tag{77}$$

Circulant. For the circulant class (Appendix D with $d = m$), let F_d be the $d\times d$ DFT matrix and define

$$\mathcal{C}_{\text{circ}} = \left\{F_d^*\,\text{diag}(e^{i\varphi_1},\dots,e^{i\varphi_d})F_m : \varphi_k\in\mathbb{R}\right\}. \tag{78}$$

The Frobenius-best projection of any $T\in O(d)$ onto this class is the diagonal-phase Procrustes solution

$$\text{proj}_{\text{circ}}(T) = F_m^*\,\text{diag}(e^{i\hat{\varphi}_1},\dots,e^{i\hat{\varphi}_d})F_d, \qquad \hat{\varphi}_k = \arg\big((F_dTF_d^*)_{kk}\big). \tag{79}$$

The zero-frequency phase is pinned to zero to preserve the constant mode, and the self-conjugate Nyquist frequency, when present, is handled according to the real-valued convention of Appendix D. In the synthetic experiment the circulant class is used as a structured subgroup of $O(d)$ for testing transport-class projection, rather than as a time-series prior.

Empirical vs. theoretical plateaus. All entries of Table 1 are normalized per fiber coordinate, i.e., divided by d. The Det. column is the lowest value of $\mathcal{L}^*/d$ in (75), evaluated at the learned transports, observed over the 5000-epoch budget, while the Theory column is $\mathcal{L}^*_{\mathcal{C}}/d$ computed on the same edge set used during training. Det. reports means $\pm$ standard deviations and Theory reports means across three seeds. Since every transport in $\mathcal{C}$ satisfies $\mathcal{L}^* \geq \mathcal{L}^*_{\mathcal{C}}$, Det. is bounded below by Theory by construction: it lies at most 0.05% above it for the circulant class and coincides with it, up to floating-point precision, for frozen identity. This confirms that restricting the transport class to $\mathcal{C} \subseteq O(d)$ constrains the per-edge restriction maps of $\Delta_{\mathcal{F}^t_{n,d}}$ in a quantitatively predictable way. We do not claim that such an arbitrary edgewise subgroup constraint automatically lifts to a smooth global connection class on the continuous bundle.

F.1.6 Hyperparameters

See Table 4.

	Spectral stability	Transport recovery
Gaussian dimension p	$\{2, 3, 4\}$ ($d \in \{3, 6, 10\}$)	4 (d=10)
sample-size grid n	$\{50, 100, 200, 400, 800\}$	$\{16, 32, 64, 128, 256\}$
reference $n_{\max}$	$\{4000, 2000, 1000\}$ at $p \in \{2, 3, 4\}$	—
graph	kNN, k=8, W_2 distance	kNN, k=8, W_2 distance
top-k eigenvalues	32	—
seeds	3	3
epochs / patience	—	5000 / 600
optimizer	—	Adam, lr $5 \cdot 10^{-3}$, batch 256
Householder reflections	—	R=16
Euler steps for P^{LC}	50	50

Table 4: Synthetic experiments: hyperparameters for spectral stability and transport recovery.

F.2 Traffic forecasting: experimental details

F.2.1 Datasets

We use two standard traffic-speed benchmarks from [69].

METR-LA. $|\mathcal{X}_n| = 207$ loop-detector sensors on the Los Angeles highway network, recording average traffic speed at 5-minute intervals. The spatial graph follows the DCRNN convention: edge weights $W_{ij} = \exp(-d_{\mathcal{M}}(x_i, x_j)^2/\sigma^2)$ with σ set to the standard deviation of the pairwise road-network distances are thresholded to retain $W_{ij} \geq \kappa$ (with $\kappa = 0.1$) and symmetrized via $W \leftarrow \max(W, W^\top)$.

PEMS-BAY. $|\mathcal{X}_n| = 325$ sensors in the San Francisco Bay Area, with the same temporal resolution and graph construction.

For both datasets, we use $T = 12$ observed time steps as input and forecast at horizons $h \in \{3, 6, 12\}$. Train/validation/test splits follow the standard $70/10/20$ chronological partition of [69].

F.2.2 Task formulation

At each forecasting instance, the input signal is

$$\mathbf{s}_{n,T} = (\mathbf{s}_{x_1}, \ldots, \mathbf{s}_{x_n}) \in C^0(\mathcal{F}^t_{n,T}; G_n), \qquad \mathbf{s}_{x_i} \in \mathbb{R}^T, \tag{80}$$

where $n = |\mathcal{X}_n|$ and $d = T$, so the network sheaf $\mathcal{F}^t_{n,T}$ has T-dimensional stalks and Laplacian $\Delta_{\mathcal{F}^t_{n,T}} \in \mathbb{R}^{nT \times nT}$. The goal is to predict future speed vectors $\mathbf{y}^{(h)} \in \mathbb{R}^{|\mathcal{X}_n|}$ at each horizon h. The prediction loss is the mean absolute forecasting error

$$\mathcal{L}_{\text{pred}} = \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{s}_{n,T}, \mathbf{y}) \in \mathcal{D}} \sum_{h \in \{3,6,12\}} \left\| \widehat{\mathbf{y}}^{(h)}(\mathbf{s}_{n,T}) - \mathbf{y}^{(h)} \right\|_1 . \tag{81}$$

For HilbNet variants with learned transports, we add the kernel regularizer of Appendix D with weight λ, giving the full training objective $\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda \mathcal{L}_{\text{kernel}-\text{reg}}$.

F.2.3 Model variants and baselines

All HilbNet variants are discretized HilbNets (Def, 5) with polynomial filters of order K and the same per-node linear readout (DCRNN convention) mapping sheaf-filtered features to per-node horizon predictions. The only architectural difference is the admissible class of edgewise transports $P_{j\to i}^{e_{ij}} \in \mathcal{C} \subseteq O(T)$, as described in Appendix D with $d = T$. The transport hypothesis classes are the same of the previous section, briefly summarized and contextualized below.

Frozen identity. $\mathcal{C}_{\mathrm{id}} = \{I_T\}$. Neighboring sensors exchange time windows without temporal alignment. The sheaf Laplacian reduces to a standard weighted graph Laplacian applied independently to each temporal coordinate, recovering a graph convolutional network.

Circulant. Each transport is a real orthogonal circulant matrix parameterized by $\lfloor (T-1)/2 \rfloor$ frequency-wise phases per edge. This encodes a time-stationary prior: the transport can advance or delay oscillatory components across an edge but cannot arbitrarily mix temporal coordinates. This is a natural inductive bias for traffic data, where congestion patterns propagate through the road network with local delays and phase shifts.

Free $O(T)$. Each transport is parameterized by a product of $R{=}8$ Householder reflections, yielding an orthogonal matrix in a strict Householder-defined subset of $O(T)$ (see Appendix D). This is the most expressive transport class but uses $\sim 10\times$ more parameters in total than the circulant variant (e.g., 119,036 vs 11,656 on METR-LA; per edge the ratio is $\sim 20\times$, since circulant uses only $\lfloor (T-1)/2 \rfloor = 5$ phases per edge at $T{=}12$).

We also compare against two non-transport baselines. A *fiber-only MLP* processes each sensor's time window independently, ignoring graph structure. A *spatiotemporal graph baseline* applies standard graph convolution to the temporally-augmented node features but does not learn sheaf transports. Finally, we include two external baselines from the literature: FC-LSTM [69] and STAEformer [71], reported as in the cited papers (parameter counts estimated from their published configurations).

F.2.4 Detailed results

Table 2 reports MAE, RMSE, and MAPE at horizons 3, 6, and 12 (mean $\pm$ std over five seeds for our experiments).

Value of graph structure. On both datasets, the frozen-identity HilbNet improves over the fiber-only MLP baseline, confirming that sheaf diffusion over the spatial graph is beneficial even without non-trivial transports.

Value of learned transports. Both the free and circulant variants consistently outperform frozen identity at all horizons on both datasets. This confirms that the sheaf structure helpw beyond the usual graph structure.

Free vs. circulant. On METR-LA, the free-$O(T)$ model achieves the best absolute accuracy at all horizons (e.g., MAE 3.938 vs. 4.059 for circulant at horizon 12), as expected from its larger hypothesis class. On PEMS-BAY, the two variants are nearly tied: free wins MAE and MAPE at horizon 12 by ~ 0.03 mph (a 2–4σ effect over five seeds), circulant wins MAE/MAPE at horizon 3, and RMSE is statistically indistinguishable at horizons 6 and 12. In both cases, the circulant model uses roughly one tenth of the transport parameters (e.g., $11{,}656$ vs. $119{,}036$ on METR-LA), making it the most parameter-efficient HilbNet variant. This supports the central pragmatic-transport message: a structured transport class encoding a physically motivated alignment prior can recover most of the benefit of unconstrained learned transports with substantially fewer degrees of freedom.

Comparison with external baselines. Our HilbNet variants are lightweight models designed to test the value of Hilbert-sheaf structure, not to compete with large-scale spatiotemporal transformers. The external baselines (FC-LSTM, STAEformer) are included for reference and use substantially more parameters and architectural components. Nevertheless, the circulant and free HilbNets outperform FC-LSTM at all horizons on both datasets while using far fewer parameters.

F.2.5 Hyperparameters

See Table 5. All experiments are run on a single H200 GPU. Hyperparameters are chosen with a sweep. All presented variants are computed using the same codebase, and we made sure they differ only in their transport parametrization.

	METR-LA	PEMS-BAY
sensors $\lvert\mathcal{X}_n\rvert$	207	325
input window T	12	12
horizons h	$\{3, 6, 12\}$	$\{3, 6, 12\}$
input feature dim	2 (speed + time-of-day)	2 (speed + time-of-day)
graph	thresh. kernel $\exp(-d^2/\sigma^2)$, κ=0.1	thresh. kernel $\exp(-d^2/\sigma^2)$, κ=0.1
sheaf-conv layers	L=2, channel widths $[2, 16, 32]$	L=2, channel widths $[2, 16, 32]$
polynomial order K per layer	$K = [2, 2]$ (HilbNet), $K = [1, 1]$ (MLP fiber baseline)	$K = [2, 2]$ (HilbNet), $K = [1, 1]$ (MLP fiber baseline)
Householder reflections	8	8
readout	per-node linear ($T_{\text{in}} \cdot F_{\text{last}} \to T_{\text{out}}$)	per-node linear ($T_{\text{in}} \cdot F_{\text{last}} \to T_{\text{out}}$)
epochs / patience	150 / 20	150 / 20
batch size	32	32
optimizer	Adam, β=$(0.9, 0.999)$	Adam, β=$(0.9, 0.999)$
learning rate	$5{\cdot}10^{-3}$ (HilbNet); $1{\cdot}10^{-3}$ (STGNN baseline)	$5{\cdot}10^{-3}$ (HilbNet); $1{\cdot}10^{-3}$ (STGNN baseline)
LR schedule	cosine annealing, $\eta_{\min}$=10^{-6}	cosine annealing, $\eta_{\min}$=10^{-6}
weight decay / dropout	0 / 0	0 / 0
gradient clipping	$\lVert g\rVert_2 \leq 5$ (DCRNN convention)	$\lVert g\rVert_2 \leq 5$ (DCRNN convention)
kernel regularizer λ	0.01	0.01
seeds	5	5

Table 5: Traffic forecasting hyperparameters.

G Mathematical Background

G.1 Hilbert Bundles

In this section, we provide relevant background on the theory of Hilbert bundles. In particular, we define the notions of Banach and Hilbert manifolds, as well as introduce the appropriate notions of connection, parallel transport, and heat flow for bundles in this setting.

G.1.1 Banach and Hilbert manifolds

To study heat kernels for smooth Hilbert bundles, we must examine manifolds modeled on generic Banach spaces. We will assume all Banach spaces and Hilbert spaces are defined over the field of real numbers $\mathbb{R}$, unless otherwise stated.

Definition 7. *A second-countable topological space $\mathcal{M}$ is a **topological Banach manifold** if there is a Banach space $\mathcal{V}$ and an atlas $\{(U_i\ ,\ \phi_i : U_i \to \mathcal{V})\}_{i\in I}$ such that the following conditions hold:*

1. *each U_i is an open subset of $\mathcal{M}$;*
2. *each $\phi_i : U_i \to \mathcal{V}$ is a homeomorphism onto an open subset of $\mathcal{V}$;*
3. *for all i, j, $\phi_i(U_i \cap U_j)$ is an open subset of $\mathcal{V}$;*
4. *the transition map $\phi_j\phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ is a homeomorphism.*

When the Banach space $\mathcal{V}$ is specified, we say that $\mathcal{M}$ is an $\mathcal{V}$-manifold, or a Banach manifold modeled on $\mathcal{V}$. If each map ϕ_i and transition map $\phi_j\phi_i^{-1}$ is k-times Fréchet differentiable, we say that $\mathcal{M}$ is a C^k-Banach manifold. If these maps are smooth i.e. C^∞, we say that $\mathcal{M}$ is a smooth Banach manifold.

Definition 8. *A topological (resp. C^k / smooth) Banach manifold $\mathcal{M}$ is a topological (resp. C^k / smooth) **Hilbert manifold** if it can be modeled on a Banach space $\mathcal{V}$ which admits the structure of a Hilbert space.*

Remark 1. *We make a few observations about this definition.*

1. *Since every n-dimensional real Banach space is isomorphic to $\mathbb{R}^n$, a finite dimensional Banach manifold is exactly a real manifold in the usual sense.*
2. *Like ordinary manifolds, we require Banach manifolds to be second countable, and hence to have a countable dense subset. It follows that if $\mathcal{M}$ is a manifold modeled on a Banach space $\mathcal{V}$, then $\mathcal{V}$ itself must be separable. This condition could be be removed, but it will generally make our lives easier.*
3. *The definition of a Hilbert manifold does not directly require the transition maps $\tau_{ij} := \phi_j \circ \phi_i^{-1} : \mathcal{V} \to \mathcal{V}$ to respect inner product structure on the modeling Hilbert space $\mathcal{V}$. Hence, it is often better to think of a Hilbert manifold as a special case of a Banach manifold, instead of as a manifold that respects the Hilbert space structure per se.*

The usual differential geometric constructions on manifolds extend naturally to Hilbert and Banach manifolds. For example, tangent spaces generalize naturally. Given a C^k-Banach manifold $\mathcal{M}$ with $k \geq 1$, for each $x \in \mathcal{M}$, one may form the **tangent space** $T_x\mathcal{M}$ at a point $x \in \mathcal{M}$ as equivalence

classes of triples (U, ϕ, v) of a chart $\phi : U \to \mathcal{V}$ and a vector $v \in \mathcal{V}$, under the relation:

$$(U_1, \phi, v) \sim (U_2, \psi, w) \iff (D_{\phi(x)}(\psi\phi)^{-1})(v) = w.$$

Such equivalence classes are easily seen to form a real vector space isomorphic to $\mathcal{V}$.

G.1.2 Smooth bundles

Definition 9 (Smooth Banach and Hilbert bundles)**.** *Let $\mathcal{M}$ be a smooth finite–dimensional manifold and let $\mathcal{V}$ be a fixed separable Banach space. A **smooth Banach bundle** with model space $\mathcal{V}$ consists of a smooth Banach manifold $\mathcal{E}$ equipped with a smooth surjective submersion*

$$\pi \colon \mathcal{E} \longrightarrow \mathcal{M}\,,$$

that satisfies the following conditions.

1. ***Local triviality.*** *For every $p \in \mathcal{M}$ there exists an open neighborhood $U \subset \mathcal{M}$ and a diffeomorphism*

   $$\phi_U \colon \pi^{-1}(U) \xrightarrow{\cong} U \times \mathcal{V}$$

   *satisfying $\pi = \mathrm{proj}_1 \circ \phi_U$, where $\mathrm{proj}_1 : U \times \mathcal{V} \to U$ is the canonical projection, and such that, for each $q \in U$, the restriction $\phi_U|_{\mathcal{E}_q} : \mathcal{E}_q \to \{q\} \times \mathcal{V}$ is a bounded linear isomorphism. We call the pair (U, ϕ_U) a **trivializing chart**.*
2. ***Smooth transition functions.*** *Whenever (U, ϕ_U) and (V, ϕ_V) are trivializing charts, the **transition map***

   $$\tau_{UV}(q, -) \;:=\; \phi_V \circ \phi_U^{-1}\big(q, -\big) \;:\; \mathcal{V} \longrightarrow \mathcal{V}, \qquad q \in U \cap V,$$

   is a bounded isomorphism and depends smoothly on q; that is, $\tau_{UV} : U \cap V \to \mathrm{GL}(\mathcal{V})$ is a smooth map, where $\mathrm{GL}(\mathcal{V})$ denotes the Banach–Lie group of bounded invertible operators on $\mathcal{V}$ with the operator-norm topology.
3. ***Smooth norm.*** *There is a smooth map $N : \mathcal{E} \to \mathbb{R}$ such that the trivializing charts (U, ϕ_U) can be chosen with the additional property that for each $x \in \mathcal{E}_p$,*

   $$N(x) = \Big\| \phi_U\big|_{\mathcal{E}_p}(x) \Big\|_{\mathcal{V}}\,.$$

4. ***Smooth fiberwise operations.*** *The fiberwise addition and scalar-multiplication maps*

   $$+ \;:\; \mathcal{E} \times_{\mathcal{M}} \mathcal{E} \longrightarrow \mathcal{E}, \qquad \cdot \;:\; \mathbb{R} \times \mathcal{E} \longrightarrow \mathcal{E},$$

   are smooth Banach-manifold maps.

*When the Banach space $\mathcal{V}$ is a separable Hilbert space, we say $\pi : \mathcal{E} \to B$ is a **Hilbert bundle**. We denote the inner product on the fiber $\mathcal{E}_p$ by $\langle -, - \rangle_p$.*

Remark 2. *We make a few remarks about the definition of a Hilbert bundle above.*

1. *For convenience, we restrict our attention to Hilbert bundles over closed finite-dimensional manifolds with separable fibers. None of these restrictions are essential for the general theory of Banach and Hilbert bundles. However, these restrictions are necessary for our approach to constructing heat kernels in this setting.*
2. *The intuitive idea is the following: a smooth Banach bundle is a smooth vector bundle where the fibers are allowed to be infinite-dimensional and come equipped with a complete norm. The smooth norm condition enforces that the Banach space fibers are stitched together in such a way that the fiber-wise norm varies smoothly. In the case of a Hilbert bundle, the smooth norm condition also enforces that the fiber-wise inner product $\langle -, - \rangle_p$ varies smoothly.*
3. *In light of the previous remarks, the definition presented here is not minimal. We make the choice to include redundant information in our definition for clarity, with the understanding that some conditions are superfluous. For instance, smooth fiberwise operations are implied by local triviality [75]. We also make the choice to include the smooth norm condition, often called a **smooth orthogonal/Hermitian metric**, in the definition of the bundle itself.*
4. *In the case of a finite dimensional model space $\mathcal{V}$, this definition recovers the usual smooth vector bundle, with the additional data of a smooth orthogonal/Hermitian metric.*

5. *Suppose $\pi : \mathcal{E} \to \mathcal{M}$ is a smooth Hilbert bundle modeled on a Hilbert space $\mathcal{H}$. While the transition maps τ_{UV} must respect the topological structure of the Hilbert space $\mathcal{H}$, it need not respect the inner product structure. When each transition map τ_{UV} is a unitary isomorphism, we say the bundle is a **smooth unitary Hilbert bundle**.*

Definition 10 (Smooth sections of a Banach bundle)**.** *Let $\pi : \mathcal{E} \to \mathcal{M}$ be a smooth Banach bundle (resp. Hilbert bundle) over a finite–dimensional manifold $\mathcal{M}$ with model Banach space (resp. Hilbert space) B. A **section** of $\mathcal{E}$ is a map $S : \mathcal{M} \to \mathcal{E}$ such that $\pi \circ S = \mathrm{id}_{\mathcal{M}}$. We denote the collection of all smooth sections by $\Gamma(\mathcal{E}) := C^\infty(\mathcal{M}, \mathcal{E})$. Note that this is a module over the commutative algebra $C^\infty(\mathcal{M})$ with point-wise addition and multiplication. If the section S is only k-times continuously differentiable, we write $S \in C^k(\mathcal{M}, \mathcal{E})$.*

Definition 11 (L^2-Sections of a Banach bundle)**.** *Suppose that the manifold $\mathcal{M}$ is endowed with a measure μ. We say that S is an L^2**-section** if $\|S\|_2 := \left(\int_{\mathcal{M}} ||S(x)||^2_{\mathcal{E}_x}\, d\mu(x)\right)^{1/2} < \infty$. We may similarly form a **space of L^2-sections**, denoted $L^2(\mathcal{M}, \mathcal{E}; \mu)$, or simply $L^2(\mathcal{M}, \mathcal{E})$ when the measure is implied by context. When $\mathcal{E}$ is modeled on a separable Hilbert space $\mathcal{H}$, the space of L^2-sections $L^2(\mathcal{M}, \mathcal{E}; \mu)$ is a real separable Hilbert space with inner product $\langle S, S' \rangle_{\mathcal{E}} := \int_{\mathcal{M}} \langle S(x), S'(x) \rangle_{\mathcal{E}_x}\, d\mu(x)$.*

Remark 3. *In the Riemannian setting, we have a natural candidate for μ via the (pseudo) volume form on $\mathcal{M}$, or its normalized variant.*

G.1.3 Connections

We now introduce connections on smooth Banach and Hilbert bundles. The smooth Banach manifold structure on the bundle $\mathcal{E}$ provides no way to directly compare vectors in different fibers $\mathcal{E}_x$ and $\mathcal{E}_y$ with respect to their Banach space structures. Instead, as in the finite-dimensional case, we use a connection to link the fibers though the geometry of the base manifold $\mathcal{M}$.

Definition 12. *Let $\pi : \mathcal{E} \to \mathcal{M}$ be a smooth Banach bundle over a compact manifold $\mathcal{M}$, and let $T^*\mathcal{M}$ denote the cotangent bundle on $\mathcal{M}$. A **connection** on $\mathcal{E}$ is any of the following three equivalent structures:*

1. *A connection is an $\mathbb{R}$-linear map:*

$$\nabla : \Gamma(\mathcal{E}) \to \Gamma(T^*\mathcal{M} \otimes \mathcal{E})$$

   such that the product rule:

$$\nabla(fS) = Df \otimes S + f \nabla S$$

   holds for all smooth function $f : \mathcal{M} \to \mathbb{R}$ and smooth sections $s \in \Gamma(\mathcal{E})$.

2. *A connection is a map $\nabla : \Gamma(T\mathcal{M}) \times \Gamma(\mathcal{E}) \to \Gamma(\mathcal{E})$ which is $C^\infty(\mathcal{M}, \mathbb{R})$-linear in its vector-field input, and satisfies the Leibniz rule:*

$$\nabla_X(fS) = X[f]S + f \nabla_X S$$

   where $\nabla_X S := \nabla(X, S)$ and $X[f](x) := (D_x f)(X_x)$ is the directional derivative of f along the vector field X.

3. *A connection ∇ is the data of an $\mathbb{R}$-linear map $\nabla_x S : T_x\mathcal{M} \to \mathcal{E}_x$ for each $x \in \mathcal{M}$ and $s \in \Gamma(\mathcal{E})$ that satisfies the following conditions:*
   (a) *$\nabla_{(-)} S$ depends smoothly on x*
   (b) *$\nabla_x(a_1 S_1 + a_2 S_2) = a_1 \nabla_x S_1 + a_2 \nabla_x S_2$ for all $x \in M$, $S_1, S_2 \in \Gamma(\mathcal{E})$, and $a_1, a_2 \in \mathbb{R}$;*
   (c) *for every smooth $f : \mathcal{M} \to \mathbb{R}$ and section $S \in \Gamma(\mathcal{E})$, let $fS \in \Gamma(\mathcal{E})$ denote the section $(fS)(x) := f(x)s(x)$. For each $x \in \mathcal{M}$, the maps $\nabla_x(fS)$ and $\nabla_x s$ are related by*

$$\nabla_x(fS)(v) = D_x f(v) S(x) + f(x)(\nabla_x S)(v)$$

   for all $v \in T_x\mathcal{M}$.

The following proposition provides a standard representation theorem for a smooth connection ∇.

Proposition 2. *Let ∇ be a smooth connection on a trivial Hilbert bundle $\pi : \mathcal{M} \times \mathcal{H} \to \mathcal{M}$. There is a map*

$$A : \Gamma(\mathcal{M} \times \mathcal{H}) \to \Gamma(T^*\mathcal{M} \otimes (\mathcal{M} \times \mathcal{H}))$$

such that for every section $S \in \Gamma(\mathcal{M} \times \mathcal{H})$, we have:

$$\nabla S = dS + AS$$

where d is the Fréchet derivative. Equivalently, for each $x \in \mathcal{M}$, and $v \in T_x M$, there is a bounded linear operator $A_{x,v} : \mathcal{H} \to \mathcal{H}$, varying smoothly in (x, v), such that:

$$(\nabla_x S)(v) = (D_x S)(v) + (A_{x,v} S)(x)\,.$$

Moreover, the assignment $v \mapsto A_{x,v}$ is linear for each x.

Remark 4. *While this proposition is stated for trivial bundles, it can be applied to any Hilbert bundle through the choice of a trivialization, or through Kuiper's theorem [61].*

G.1.4 Parallel Transport

Connections allow us to relate the geometries of the fibers over nearby points in $\mathcal{M}$. For example, we may define parallel transport.

Given a smooth curve $\gamma : [0, 1] \to \mathcal{M}$, say that $S : [0, 1] \to \mathcal{E}$ is a **section over** γ if $\pi \circ S = \gamma$.

Definition 13. *Let $\pi : \mathcal{E} \to \mathcal{M}$ be a smooth Banach bundle with model space X, let ∇ be a connection on $\mathcal{E}$, and let $\gamma : [0, 1] \to \mathcal{M}$ be a smooth path. A map $S : [0, 1] \to \mathcal{E}$ is a **section over** γ if $\pi \circ S = \gamma$. A section over γ is **parallel** if*

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

Proposition 3. *Let γ be a smooth path in $\mathcal{M}$. For every vector $v \in \mathcal{E}_{\gamma(0)}$, there is a unique parallel section S_v over γ such that $S_v(0) = v$. Moreover, the dependence on v is smooth and linear.*

Proof. The existence of the parallel section S_v can be restated as an initial value problem for a linear ordinary differential equation in a Banach space. Standard existence and uniqueness theorems apply. For details, see [62]. □

By the existence and uniqueness of parallel sections, we may define corresponding **parallel transport maps**. Given a path $\gamma : [0, 1] \to \mathcal{M}$, there is an induced **parallel transport** operator $P^\nabla_\gamma : \mathcal{E}_{\gamma(0)} \to \mathcal{E}_{\gamma(1)}$ defined by

$$P^\nabla_\gamma(v) := S_v(1).$$

It is straightforward to see that P^∇_γ is a linear bijection, with inverse given by $(P^\nabla_\gamma)^{-1} = P^\nabla_{\gamma^{\text{rev}}}$, where γ^{rev} is the path obtained by reversing γ. By the closed graph theorem, it follows that P^∇_γ is a bounded linear isomorphism.

Definition 14. *Let $\pi : \mathcal{E} \to \mathcal{M}$ be a Hilbert bundle equipped with a connection ∇. We say the connection ∇ is **compatible** with the Hilbert bundle structure if it satisfies:*

$$X[\langle S_0(x), S_1(x)\rangle_x] = \langle \nabla_X S_0(x), S_1(x)\rangle_x + \langle S_0(x), \nabla_X S_1(x)\rangle_x$$

for every smooth vector field $X \in \Gamma(T\mathcal{M})$, and sections $S_0, S_1 \in \Gamma(\mathcal{E})$.

Proposition 4. *Let $\pi : \mathcal{E} \to \mathcal{M}$ be a Hilbert bundle with connection ∇. The following are equivalent:*

1. *∇ is compatible with the Hilbert bundle structure;*
2. *Every parallel transport map P^∇_γ is unitary;*

Proof. First suppose that ∇ is compatible with the Hilbert bundle structure, and γ a smooth path in $\mathcal{M}$. Let $u, v \in \mathcal{E}_{\gamma(0)}$. By compatibility, we may check that:

$$\frac{d}{dt}\big\langle S_u(t), S_v(t)\big\rangle_{\gamma(t)} = \big\langle \nabla_{\dot\gamma} S_u(t), S_v(t)\big\rangle_{\gamma(t)} + \big\langle S_u(t), \nabla_{\dot\gamma} S_v(t)\big\rangle_{\gamma(t)} = 0\,.$$

It immediately follows that P^∇_γ is unitary.

Conversely, suppose that every parallel transport map P^∇_γ is unitary. Let X be a smooth vector field, and S_0, S_1 smooth sections of $\mathcal{E}$. Let $x \in \mathcal{M}$, and let γ be a smooth path such that $\gamma(0) = x$. For $j \in \{0, 1\}$ let u_j be a parallel section over γ such that $u_j(0) := S_j(x)$. For $j \in \{0, 1\}$, let

$w_j(t) := S_j(\gamma(t)) - u_j(t)$. Finally, let $w_j(t) := S_j(\gamma(t)) - u_j(t)$. Since u_j is parallel over γ, we have that $\nabla_{\dot\gamma} S_j(\gamma(t)) = \nabla_{\dot\gamma} w_j(t)$. Moreover, $w_j(0) = 0$. We may use these facts to compute:

$$\begin{aligned} X[\langle S_0(x), S_1(x)\rangle_x] &= \frac{d}{dt}\Big|_{t=0} \langle S_0(\gamma(t)), S_1(\gamma(t))\rangle_{\gamma(t)} \\ &= \frac{d}{dt}\Big|_{t=0} \langle u_0(t) + w_0(t), u_1(t) + w_1(t)\rangle_{\gamma(t)} \\ &= \frac{d}{dt}\Big|_{t=0} \langle u_0(t), u_1(t)\rangle_{\gamma(t)} + \langle \nabla_{\dot\gamma} S_0(x), S_1(x)\rangle_x + \langle S_0(x), \nabla_{\dot\gamma} S_1(x)\rangle_x \end{aligned}$$

Since parallel transport maps are unitary, it follows that the quantity $\langle u_0(t), u_1(t)\rangle_{\gamma(t)}$ is constant in t. Hence

$$\begin{aligned} X[\langle S_0(x), S_1(x)\rangle_x] &= \frac{d}{dt}\Big|_{t=0} \langle u_0(t), u_1(t)\rangle_{\gamma(t)} + \langle \nabla_{\dot\gamma} S_0(x), S_1(x)\rangle_x + \langle S_0(x), \nabla_{\dot\gamma} S_1(x)\rangle_x \\ &= \langle \nabla_X S_0(x), S_1(x)\rangle_x + \langle S_0(x), \nabla_X S_1(x)\rangle_x \end{aligned}$$

proving that ∇ is a compatible connection. □

Remark 5. *We note that by the Proposition 4, as we assumed our parallel transport operators to be unitary in the main text, this may equivalently be understood as a metric-compatibility condition.*

G.2 Connection Laplacian

Let $\pi : \mathcal{E} \to \mathcal{M}$ be a Hilbert bundle on a closed Riemannian manifold, equipped with a compatible connection $\nabla : \Gamma(\mathcal{M}; \mathcal{E}) \to \Gamma(\mathcal{M}; T^*\mathcal{M} \otimes \mathcal{E})$. Moreover, $T^*\mathcal{M} \otimes \mathcal{E}$ inherits the structure of a Hilbert bundle, with fiber-wise inner products induced from the metric g and the fiber-wise inner products of $\mathcal{E}$. Since ∇ is a linear differential operator, it has a formal adjoint $\nabla^* : \Gamma(\mathcal{M}; T^*\mathcal{M} \otimes \mathcal{E}) \to \Gamma(\mathcal{M}; \mathcal{E})$, defined implicitly by the formula:

$$\int_{\mathcal{M}} \langle \nabla S_0(x), S_1(x)\rangle_x \, d\mu(x) \;=\; \int_{\mathcal{M}} \langle S_0(x), \nabla^* S_1(x)\rangle_x \, d\mu(x)$$

where $S_0 \in \Gamma(\mathcal{M}; \mathcal{E})$, $S_1 \in \Gamma(\mathcal{M}; T^*\mathcal{M} \otimes \mathcal{E})$, and μ is the pseudo-volume form on $\mathcal{M}$. Using this adjoint, we may define the connection Laplacian.

Definition 15. *The **connection Laplacian** is the linear operator:*

$$\Delta_\nabla := \nabla^*\nabla : \Gamma(\mathcal{E}) \to \Gamma(\mathcal{E}) \,.$$

Remark 6. *The connection $\nabla : \Gamma(\mathcal{E}) \to \Gamma(T^*\mathcal{M} \otimes \mathcal{E})$ can be extended to a closed, densely-defined unbounded operator $\nabla_{L^2} : L^2(\mathcal{E}) \to L^2(T^*\mathcal{M} \otimes \mathcal{E})$. This extended operator has an adjoint $\nabla^*_{L^2}$ as a Hilbert space operator, from which we may define a composite $\Delta_{\nabla_{L^2}} := \nabla^*_{L^2}\nabla_{L^2}$. The formal adjoint ∇^* and connection Laplacian Δ_∇ can be found by restricting the domains of $\nabla^*_{L^2}$ and $\Delta_{\nabla_{L^2}}$ to linear subspaces of smooth sections. From this perspective, it becomes clear that the connection Laplacian Δ_∇ is well defined for all C^2 sections, as C^2 is contained in the Sobolev space H^2.*

The connection Laplacian also admits a characterization in terms of covariant derivatives. Let $\nabla^2_{X,Y}$ denote the second covariant derivative with respect to vector fields X, Y.

Lemma 1. *(Connection Laplacian in Coordinates) Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle over a closed Riemannian manifold equipped with a compatible Fréchet connection. As operators,*

$$\Delta_\nabla S = -\mathrm{tr}\nabla^2 S \,.$$

Moreover, with respect to a local orthonormal frame $\{e_i\}_{i=1}^m$ that is synchronous at $p \in \mathcal{M}$, we have equality:

$$\Delta_\nabla S(p) = -\sum_{i=1}^m \nabla_{e_i}\nabla_{e_i} S(p) \,.$$

Proof. We adapt the proof of [79] to the Hilbert bundle setting, using the synchronous frame technique of [76]. By a partition of unity argument, it suffices to show that for every pair of smooth sections S_1, S_2 supported inside the domain of a local orthonormal frame $\{e_i\}_{i=1}^m$, we have an equality:

$$\int_{\mathcal{M}} \langle \Delta_\nabla S_1(x), S_2(x)\rangle_x \, d\mu(x) = -\int_{\mathcal{M}} \sum_{i=1}^m \langle \nabla^2_{e_i,e_i} S_1(x), S_2(x)\rangle_x \, d\mu(x) \,.$$

Using the formal adjoint of ∇, the integral on the left can be rewritten as

$$\int_{\mathcal{M}} \langle \Delta_\nabla S_1(x), S_2(x)\rangle_x \, d\mu(x) = \sum_{i=1}^m \int_{\mathcal{M}} \langle \nabla_{e_i} S_1(x), \nabla_{e_i} S_2(x)\rangle_x \, d\mu(x) \, .$$

To analyze the right hand side, simply note that $\nabla^2_{e_i,e_i} S_1 = \nabla_{e_i}\nabla_{e_i} S_1 - \nabla_{\nabla_{e_i} e_i} S_1$, and by compatibility, that:

$$e_i[\langle \nabla_{e_i} S_1(x), S_2(x)\rangle_x] = \langle \nabla_{e_i}\nabla_{e_i} S_1(x), S_2(x)\rangle_x + \langle \nabla_{e_i} S_1(x), \nabla_{e_i} S_2(x)\rangle_x \, .$$

Rearranging and summing over i yields:

$$\begin{aligned}\sum_i \langle \nabla^2_{e_i,e_i} S_1(x), S_2(x)\rangle_x = &- \sum_i \langle \nabla_{e_i} S_1(x), \nabla_{e_i} S_2(x)\rangle_x \\ &+ \sum_i \left(e_i[\langle \nabla_{e_i} S_1(x), S_2(x)\rangle_x] - \langle \nabla_{\nabla_{e_i} e_i} S_1(x), S_2(x)\rangle_x \right) \, .\end{aligned}$$

The second sum on the right-hand side may be identified as the divergence of the vector field $\langle \nabla_v S_1(x), S_2(x)\rangle_x$, and hence integrates to zero by Stokes' theorem. Therefore

$$-\int_{\mathcal{M}} \sum_i \langle \nabla^2_{e_i,e_i} S_1(x), S_2(x)\rangle_x \, d\mu(x) = \sum_i \int_{\mathcal{M}} \langle \nabla_{e_i} S_1(x), \nabla_{e_i} S_2(x)\rangle_x \, d\mu(x)$$

as well. Therefore $\Delta_\nabla(S) = -\mathrm{tr}\nabla^2(S)$. Under the additional hypothesis that $\{e_i\}_{i=1}^m$ is synchronous at p, we have $\nabla_{e_i} e_i = 0$ for each $1 \le i \le m$. At such a point, the trace reduces to $\mathrm{tr}\nabla^2 S(p) = \sum_i \nabla_{e_i}\nabla_{e_i} S(p)$. □

G.3 Heat Flow on a Hilbert Bundle

In this section, we fix a closed finite-dimensional Riemannian manifold $(\mathcal{M}, g)$ with canonical volume pseudo-form μ, a smooth Hilbert bundle $\pi : \mathcal{E} \to \mathcal{M}$ with fiber $\mathcal{E}_x \cong \mathcal{H}$, and a connection ∇. Our goal in this section is three-fold:

1. Demonstrate that the heat equation with respect to the connection Laplacian Δ_∇ has a unique solution;
2. Show that the heat-flow admits a Heat Kernel;
3. Provide asymptotic estimates for the heat flow that relate to the geometry of the underlying manifold $\mathcal{M}$.

Our approach is to adapt the methods of Berline, Getzler, and Vergne [15] for finite-rank bundles to the Hilbert bundle setting. A key subtlety that arises in this generalization is in the definition of tensor-products of bundles. While the algebraic and topological tensor products of finite-rank Hilbert spaces agree, they need not coincide for general Hilbert spaces. This complicates, for instance, the necessary tensor-hom adjunction. In order to keep track of the appropriate tensor, we adopt the following convention. Let $\pi_0 : \mathcal{E}_0 \to \mathcal{M}$ and $\pi_1 : \mathcal{E}_1 \to \mathcal{M}$ be smooth Hilbert bundles over a common manifold $\mathcal{M}$. One may form the **hom-bundle** $\hom(\mathcal{E}_0, \mathcal{E}_1) \to \mathcal{M}$, whose fiber $\hom(\mathcal{E}_0, \mathcal{E}_1)_x$ is the Banach space of bounded linear operators from $\mathcal{E}_0$ to $\mathcal{E}_1$. This is a Banach bundle when $\hom(\mathcal{E}_0, \mathcal{E}_1)_x$ is topologized with the operator norm.

Definition 16. *Let Δ_∇ be the connection Laplacian for a compatible connection on a smooth Hilbert bundle $\pi : \mathcal{E} \to \mathcal{M}$ over a compact orientable manifold $\mathcal{M}$. A **heat kernel** for Δ_∇ is a continuous section $K^\nabla(x, y, t)$ of the Banach bundle $\hom\big(pr_2^*\mathcal{E}, pr_1^*\mathcal{E}\big)$ over $\mathcal{M} \times \mathcal{M} \times (0, \infty)$ that satisfies the following conditions.*

1. *$K^\nabla(x, y, t)$ is C^1 with respect to t, and is C^2 with respect to x.*
2. *$K^\nabla(x, y, t)$ satisfies the **heat equation** $\partial_t K^\nabla(x, y, t) = -(\Delta_\nabla)_x K^\nabla(x, y, t)$, where $(\Delta_\nabla)_x$ means applying the Laplacian to x.*
3. *$K^\nabla(x, y, t)$ satisfies the boundary condition $\lim_{t\to 0} K_t S = S$ for every smooth section $s \in \Gamma(\mathcal{E})$, where $(K_t S)(x) = \int K^\nabla(x, y, t) S(y) \, d\mu(y)$.*

Lemma 2. *(Heat Kernel for Hilbert Bundles) Let Δ_∇ be the connection Laplacian associated to a Hilbert bundle, $(\mathcal{M}, \mathcal{E}, \nabla)$. Let $n := \dim(\mathcal{M})/2$, ψ be a cutoff function, and*

$$C_n(x, y) := \frac{\exp\left(\frac{-d_{\mathcal{M}}(x,y)^2}{4t}\right)}{(4\pi t)^{n/2}}.$$

The following hold:

1. *The Laplacian Δ_∇ admits a unique heat kernel $K^\nabla(x, y, t)$.*
2. *There exist smooth sections $\Phi_i \in \Gamma\left(\mathcal{M} \times \mathcal{M}, hom\big(pr_2^*\mathcal{E}, pr_1^*\mathcal{E}\big)\right)$ such that for every $N > n$, the kernel*

$$K^N(x, y, t) := C_n(x, y)\psi(d_{\mathcal{M}}(x, y)^2) \sum_{i=0}^{N} t^i \Phi_i(x, y) j(x, y)^{-1/2} |dx|^{1/2}$$

is asymptotic to $K^\nabla(x, y, t)$, in the sense that

$$\|\partial_t^k (K^\nabla(x, y, t) - K^N(x, y, t))\|_\ell = O(t^{N - \frac{n+\ell+2k}{2}}).$$

3. *The leading term $\Phi_0(x, y)$ is equal to the parallel transport $P_{y\to x} : \mathcal{E}_y \to \mathcal{E}_x$ with respect to the Fréchet connection associated to $\mathcal{H}$ along the unique length-minimizing geodesic joining y and x.*

Proof. (Sketch) We note that parametrix-based approach of Berline, Getzler, and Vergne [15] extends to our setting with minor but judicious modifications. As the original parametrix argument is quite lengthy, we simply make note of the necessary modifications to their argument. First, all integration must be understood as Bochner integration. Second, to avoid ambiguities surrounding the algebraic and geometric tensor bundle, the hom-bundle is used instead of tensor bundles.
Now note that the parametrix argument is fundamentally local in nature. Consider a smooth Hilbert bundle $\mathcal{E} \to \mathcal{M}$, and note that we require an associated Fréchet connection ∇ to be metric-compatible. Then, in a a local trivialization $\mathcal{E}\big|_U \cong U \times \mathcal{H}$, with $\mathcal{H}$ a separable Hilbert space, the connection has the form $\Delta = d + A$ with A a smooth $\hom(\mathcal{H}, \mathcal{H})$ valued 1-form by Proposition 2. The connection Laplacian is a second-order elliptic operator with scalar principal symbol and lower-order coefficients in the Banach algebra $\hom(\mathcal{H}, \mathcal{H})$. The parametrix argument then proceeds via solving transport equations along geodesics and then correcting the resulting approximated kernel by a Volterra series. These steps do not rely in any essential way on finite-dimensionality of the fiber, but only on the fact that the coefficient algebra admits the usual smooth calculus and operator-norm estimates. Thus, replacing matrix-valued coefficients by $\hom(\mathcal{H}, \mathcal{H})$-valued ones, one obtains in the same way a smooth kernel $K^\nabla(x, y, t)$, and the usual energy argument gives uniqueness. □

Remark 7. *The details of the necessary parametrix argument may be found in 2.1 – 2.5 of [15]. Note that while they assume a finite-rank hypothesis through the entirety of chapter two, the hypothesis is actually unused until section 2.6 of their work, when the operator K_t is required to be Hilbert-Schmidt.*

G.4 Borel Functional Calculus

Given a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ and a suitably well-behaved function $g : \mathbb{R} \to \mathbb{C}$, one may "apply g" to T to get a new linear map $g(T) : \mathbb{R}^n \to \mathbb{R}^n$. In particular, whenever g is analytic with a globally defined Maclaurin expansion $g(x) = \sum_j a_j x^j$, one may define $g(T) = \sum_j a_j T^j$, where T^j is interpreted as the j-fold composition $T \circ \cdots \circ T$. When T is a bounded linear endo-operator on a Hilbert space, one may similarly define $g(T)$ via series expansion. However, when T is unbounded, more care must be taken to handle series convergence. This difficulty in the unbounded case is pertinent for the HilbNet architecture, where the convolution filter g must be applied to the unbounded connection Laplacian Δ_∇. The *Borel functional calculus* provides an elegant solution. While traditionally formulated through the spectral theorem and projection-valued measures, for the purpose of the HilbNet architecture, the following version (Theorem VIII.5 of [82]) will be sufficient.

Theorem 3. *(Spectral Theorem - Functional Calculus Form) Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Then there is a unique map $\hat{\phi}$ from the bounded Borel measurable functions on $\mathbb{R}$ into the space of bounded linear operators on $\mathcal{H}$, $\mathcal{L}(\mathcal{H})$, so that*

- *$\hat{\phi}$ is an algebraic *-homomorphism.*
- *$\hat{\phi}$ is norm continuous, that is, $\|\hat{\phi}(h)\|_{\mathcal{L}(\mathcal{H})} \leq \|h\|_\infty$.*
- *Let $g_n(x)$ be a sequence of bounded Borel functions with $g_n(x) \xrightarrow[n\to\infty]{} x$ for each x and $|h_n(x)| \leq |x|$ for all x and n. Then, for any $\psi \in \mathrm{dom}(A)$, $\lim_{n\to\infty} \hat{\phi}(h_n)\,\psi = A\psi$.*

- *If $g_n(x) \to h(x)$ pointwise and if the sequence $\|h_n\|_\infty$ is bounded, then $\hat{\phi}(h_n) \to \hat{\phi}(h)$ strongly.*

In addition:

- *If $A\psi = \lambda\psi$, then $\hat{\phi}(h)\psi = h(\lambda)\psi$.*
- *If $g \geq 0$, then $\hat{\phi}(h) \geq 0$.*

G.5 Cellular Sheaves and Sheaf Laplacians

Cellular sheaves on graphs are a data structure that generalizes weighted graphs. We take our exposition of cellular sheaves and their Laplacians primarily from [51]. See [51, 29, 53] for more details.

Definition 17 (Cellular Sheaf on a Graph)**.** *Let $G = (V, E)$ be an undirected multi-graph without self-loops, and finitely many vertices and edges. Let $v \leq e$ denote that node v is incident to the edge e. A **cellular sheaf**, or equivalently **network sheaf**, $\mathcal{F}$ on G consists of the following data.*

- *A vector space $\mathcal{F}(\sigma)$ for each $\sigma \in V \amalg E$, called the **stalk** over σ.*
- *A linear map $\mathcal{F}_{v \leq e} : \mathcal{F}(v) \to \mathcal{F}(e)$ for each incident pair $v \leq e$, called the **restriction map** of v into e.*

Remark 8. *At the level of category theory, a cellular sheaf is a functor $\mathcal{F} : G \to \mathbf{Vect}$, where the graph $G = (V, E)$ is viewed as a posetal category with objects $\mathrm{Ob}(G) = V \amalg E$, and a unique homomorphism from $v \to e$ whenever $v \leq e$. In this light, we adopt the notation $\mathcal{F} : G \to \mathbf{Vect}$ for a cellular sheaf $\mathcal{F}$ on a graph G.*

Traditionally, to add geometric content to a cellular sheaf, one passes to *weighted cellular sheaves*: a cellular sheaf $\mathcal{F} : G \to \mathbf{Vect}$ where each stalk $\mathcal{F}_\sigma$ is a finite dimensional vector space endowed with an inner product $\langle -, - \rangle_\sigma$. To accommodate infinite-dimensional Hilbert space stalks, we instead follow the approach of [44].

Definition 18. *(Hilbert Cellular Sheaf on a Graph) A **Hilbert cellular sheaf** $\mathcal{F}$ on a finite graph $G = (V, E)$ consists of the following data.*

- *A Hilbert space $\mathcal{F}(v)$ for each $v \in V$, referred to as the **node stalk** over v.*
- *A Hilbert space $\mathcal{F}(e)$ for each $e \in E$, referred to as the **edge stalk** over v.*
- *For each edge $e \in E$ with bounding vertices u, v, a pair of bounded linear **restriction maps** $\mathcal{F}_{u \leq e} : \mathcal{F}(u) \to \mathcal{F}(e)$ and $\mathcal{F}_{v \leq e} : \mathcal{F}(v) \to \mathcal{F}(e)$.*

Remark 9. *A bounded Hilbert sheaf $\mathcal{F}$ can again be viewed as a functor $\mathcal{F} : G \to \mathbf{Hilb}_\mathbb{R}$, where G is the graph G viewed as an acyclic category, and $\mathbf{Hilb}_\mathbb{R}$ is the category of real Hilbert spaces and bounded globally-defined linear operators.*

Remark 10. *In order to better differentiate between the usual finite-rank cellular sheaf on a graph and the potentially infinite-rank Hilbert cellular sheaves considered above, we use the terminology of network sheaves when we wish to emphasize the finite-rank consideration.*

Definition 19. *Let $\mathcal{F} : G \to \mathbf{Hilb}_\mathbb{R}$ be a bounded Hilbert sheaf on a graph $G = (V, E)$. The spaces of **0-cochains** and **1-cochains** are defined by:*

$$C^0(G; \mathcal{F}) := \bigoplus_{v \in V} \mathcal{F}(v)\,,$$
$$C^1(G; \mathcal{F}) := \bigoplus_{e \in E} \mathcal{F}(e)\,.$$

where $\oplus$ denotes the direct sum of Hilbert spaces. For a 0-cochain $\mathbf{x} \in C^0(G; \mathcal{F})$, we denote the component of $\mathbf{x}$ in the stalk over the node v by $\mathbf{x}_v$, with a similar notation for components of 1-cochains.

Definition 20. *Let $G = (V, E)$ be a graph. A **signed incidence relation** on G is a pairing $[- : -] : V \times E \to \{-1, 0, 1\}$ which satisfies the following conditions:*

1. *$[v : e] \neq 0$ if and only if $v \leq e$.*
2. *For each e, $\sum_{v \leq e} [v : e] = 0$.*

Remark 11. *The data of a signed incidence structure on $G = (V, E)$ is equivalent to the choice of a source $s(e)$ and target $t(e)$ for each edge e. In particular, the total set of incidences can be put into two-to-one correspondence with edges, counting the two distinct "boundings" of each $e \in E$.*

Definition 21. *Let $G = (V, E)$ be a graph equipped with a signed incidence relation. Let $\mathcal{F} : G \to \mathbf{Hilb}_{\mathbb{R}}$ be a bounded Hilbert sheaf on G. The **coboundary operator** $\delta : C^0(G; \mathcal{F}) \to C^1(G; \mathcal{F})$ is the operator with image on the each edge stalk:*

$$(\delta \mathbf{x})_e := \sum_{v \leq e} [v : e] \mathcal{F}_{v \leq e}(\mathbf{x}_v) \, .$$

Remark 12. *The coboundary map δ depends on the choice of the signed incidence relation. However, given two signed incidence relations $[- : -]_1, [- : -]_2$, the corresponding coboundary operators δ_1, δ_2 differ on the stalk $\mathcal{F}(e)$ by at most a sign difference $(\delta_1 \mathbf{x})_e = \pm(\delta_2 \mathbf{x})_e$. In particular, $\ker(\delta)$ does not depend on the choice of ϵ.*

Definition 22. *Let $\mathcal{F} : G \to \mathbf{Hilb}_{\mathbb{R}}$ be a bounded Hilbert network sheaf on a graph $G = (V, E, \epsilon)$ equipped with a signed incidence relation. Let $\delta^* : C^1(G; \mathcal{F}) \to C^0(G; \mathcal{F})$ denote the linear adjoint of the corresponding coboundary operator with respect to the inner product structures on the spaces of cochains as product spaces. The **Hilbert sheaf Laplacian** is the operator $\Delta_{\mathcal{F}} : C^0(G; \mathcal{F}) \to C^0(G; \mathcal{F})$ defined by the composition:*

$$\Delta_{\mathcal{F}} = \delta^* \circ \delta \, .$$

Proposition 5. *The Hilbert sheaf Laplacian $\Delta_{\mathcal{F}}$ has the following properties.*

1. *The Laplacian $\Delta_{\mathcal{F}}$ is a self-adjoint globally-defined bounded linear operator.*
2. *When $C^0(G; \mathcal{F}) \xrightarrow{\delta} C^1(G; \mathcal{F})$ is viewed as a Hilbert complex (in the sense of [23]) the kernel $\ker(\Delta_{\mathcal{F}})$ recovers the space of harmonic 0-cochains.*
3. *The negative Laplacian $-\Delta_{\mathcal{F}}$ is the infinitesimal generator of a strongly continuous semi-group $e^{-t\Delta_{\mathcal{F}}}$ on $C^0(G; \mathcal{F})$. For a choice of initial cochain $\mathbf{x}_0 \in C^0(G; \mathcal{F})$, the resulting flow $\mathbf{x}_t := e^{-t\Delta_{\mathcal{F}}} \mathbf{x}_0$ is a solution to the sheaf heat equation $\frac{d}{dt}\mathbf{x}_t = -\Delta_{\mathcal{F}} \mathbf{x}_t$. Moreover, the flow has limiting behavior $\lim_{t \to \infty} \mathbf{x}_t = \Pi_{\ker(\Delta_{\mathcal{F}})} \mathbf{x}_0$, where $\Pi_{\ker(\Delta_{\mathcal{F}})}$ denotes the orthogonal projection onto $\ker(\Delta_{\mathcal{F}})$.*

Proof. See [44]. □

We now recall our construction of Hilbert cellular sheaf from a spatially-discretized Hilbert bundle.

Definition (Hilbert Cellular Sheaf from a Hilbert Bundle)**.** *For a given Hilbert bundle $(\mathcal{M}, \mathcal{E}, \nabla)$ with sampled points $\mathcal{X}_n = \{x_1, \ldots, x_n\} \subset \mathcal{M}$, fix a geodesic γ_{ij} between x_i and x_j, for all $i < j$. Further, let $m_{\gamma_{ij}}$ denote the midpoint of this geodesic. Consider the graph $G_n = (\mathcal{X}_n, E)$ with an undirected edge e_{ij} between x_i and x_j, for each $i < j$. The associated **Hilbert cellular sheaf** $\mathcal{F}_n^t$ on G_n with bandwidth parameter t is given by the following assignments:*

- *The Hilbert space $\mathcal{F}_n^t(x_i) := \mathcal{E}_{x_i}$ for each $x_i \in \mathcal{X}_n$, referred to as the node stalk over $x_i \in \mathcal{X}_n$.*
- *The Hilbert space $\mathcal{F}_n^t(e_{ij}) := \mathcal{E}_{m_{\gamma_{ij}}}$ for each $e_{ij} \in E$, referred to as the edge stalk over $e_{i,j} \in E$.*
- *For each edge $e_{ij} \in E$ with bounding vertices x_i, x_j, a pair of bounded linear restriction maps*

$$\begin{aligned} (\mathcal{F}_n^t)_{x_i \leq e_{ij}} &:= \sqrt{k_{ij}^t} \, P_{x_i \to m_{\gamma_{ij}}} : \mathcal{F}_n^t(x_i) \to \mathcal{F}_n^t(e_{ij}), \\ (\mathcal{F}_n^t)_{x_j \leq e_{ij}} &:= \sqrt{k_{ij}^t} \, P_{x_j \to m_{\gamma_{ij}}} : \mathcal{F}_n^t(x_j) \to \mathcal{F}_n^t(e_{ij}), \end{aligned} \tag{82}$$

where $k_{ij}^t = e^{-d_{\mathcal{M}}(x_i, x_j)^2 / 4t}$, with $d_{\mathcal{M}}$ the geodesic distance on $\mathcal{M}$, and $P_{x_i \to m_{\gamma_{ij}}}$ denotes the unitary parallel transport map on $\mathcal{E}$ between x_i and $m_{\gamma_{ij}}$.

Remark 13. *We make a few clarifying remarks on this construction.*

- *Note that geodesics exist in this setting by the Hopf-Rinow theorem [36] by compactness of $\mathcal{M}$, and we should further choose length-minimizing geodesics.*
- *For simplicity, we use the geodesic distance to weight our restriction maps. However, we could also use the Euclidean heat kernel ala [14], and this would result in a reweighted sheaf Laplacian but would ultimately converge to the same connection Laplacian. In practical implementations, it is thus well-justified to work with the Euclidean heat kernel rather than geodesic distance based weights.*

- *While this particular construction is chosen to emphasize the relationship to [14] and allow for the necessary analytical arguments, there exist alternative constructions that are geodesic choice-independent that generate the same sheaf Laplacian but emphasize functoriality.*

G.6 Empirical Laplacians

Analogously to [14], we introduce two intermediary notions of Laplacian that interpolate between the Hilbert sheaf Laplacian and the Laplacian on a Hilbert bundle. For this section, fix a Hilbert bundle with compatible Fréchet connection $(\mathcal{M}, \mathcal{E}, \nabla)$ over a closed manifold $\mathcal{M}$.

Definition 23. *Consider the unique normalized volume pseudo-form on $\mathcal{M}$ (or the usual volume form if $\mathcal{M}$ is orientable), denoted $d\mu$. Thus, $d\mu$ equips $\mathcal{M}$ with a probability measure and we may refer to the resulting distribution as the **uniform distribution** on $\mathcal{M}$.*

Henceforth, let $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ denote the realization of an iid random sample drawn from the uniform distribution on $\mathcal{M}$. We then recall the following construction.

Definition. *(Point-Cloud Extension of Sheaf Laplacian) Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle and consider a sample $\mathcal{X}_n \subset \mathcal{M}$. Then the corresponding Hilbert sheaf Laplacian $\Delta_{\mathcal{F}_n^t}$ may be extended to the **point-cloud Laplacian** $\hat{\Delta}_{\mathcal{F}_n^t}$, an operator on $L^2(\mathcal{M}, \mathcal{E})$ via*

$$(\hat{\Delta}_{\mathcal{F}_n^t} S)(x) := \frac{1}{n} \sum_j e^{-d_{\mathcal{M}}(x, x_j)^2/4t} \big(S(x) - P_{x_j \to x} S(x_j)\big) \tag{83}$$

Remark 14. *We make the following remarks about the point-cloud Laplacian.*

1. *The point-cloud Laplacian is the extension of the Hilbert sheaf Laplacian $\Delta_{\mathcal{F}_n^t} : C^0(G; \mathcal{F}_n^t) \to C^0(G; \mathcal{F}_n^t)$ to an operator acting on sections of the Hilbert bundle $\mathcal{E} \to \mathcal{M}$, normalized by a factor of $1/n$. In particular, when evaluated at a sample point $x_i \in \mathcal{X}$, the point cloud Laplacian $\hat{\Delta}_{\mathcal{F}_n^t} S(x_i)$ is exactly the normalized x_i component of the Hilbert sheaf Laplacian $\Delta_{\mathcal{F}_n^t}$ evaluated at the cochain $(S(x_1), \ldots, S(x_n))^T \in C^0(G; \mathcal{F}_n^t)$.*
2. *The point-cloud Laplacian is well defined for any section $S : \mathcal{M} \to \mathcal{E}$, regardless of regularity.*

Definition 24 (Functional Approximation Laplacian)**.** *For a section $S \in C^1(\mathcal{E})$, we define the **functional approximation** to the connection Laplacian*

$$(\hat{\Delta}^t S)(x) := \int_{\mathcal{M}} \big(S(x) - P_{y \to x} S(y)\big) \exp\left(-\frac{d_{\mathcal{M}}(x, y)^2}{4t}\right) dy$$

where $\int(-)\, dy$ denotes Bochner integration with respect to the canonical normalized volume pseudo-form on $\mathcal{M}$.

Remark 15. *We make the following remarks about the functional approximation Laplacian.*

1. *The functional approximation Laplacian has no dependence on a sample of points from the underlying manifold. Instead, the functional approximation may be treated as the limiting operator where all points on the manifold have been sampled, and contribute uniformly via parallel transport.*
2. *The geometric data of the connection ∇ impacts the functional approximation Laplacian through the parallel transport maps $P_{y \to x} : \mathcal{E}_y \to \mathcal{E}_x$, which links the fibers of $\mathcal{E}$.*
3. *Viewing the sample $\mathcal{X}_n = \{x_1, \ldots, x_n\} \subseteq \mathcal{M}$ as having been drawn iid from the uniform probability distribution on $\mathcal{M}$, the functional approximation Laplacian can be identified pointwise on a section S as the expected value of the point cloud Laplacian. That is, for any $S \in C^1(\mathcal{M}, \mathcal{E})$ and $x \in \mathcal{M}$, we have:*

$$\frac{1}{\mathrm{vol}(\mathcal{M})} (\hat{\Delta}^t S)(x) = \mathbb{E}_{\mathcal{X}} \left[(\hat{\Delta}_{\mathcal{F}_n^t} S)(x) \right]$$

H Proofs of Results

H.1 Auxiliary Lemmas for Theorem 1

Lemma 3. *For $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, $k \in \mathbb{N}$, and $a, t > 0$, the following Gaussian identities hold:*

$$\frac{1}{(2\pi at)^{m/2}} \int x_i \exp\left(-\frac{\|x\|^2}{2at}\right) dx = 0, \tag{84}$$

$$\frac{1}{(2\pi at)^{m/2}} \int x_i x_j \exp\left(-\frac{\|x\|^2}{2at}\right) dx = at\delta_{ij}, \tag{85}$$

$$\frac{1}{(2\pi at)^{m/2}} \int \|x\|^{2k+1} \exp\left(-\frac{\|x\|^2}{2at}\right) dx = O(t^{k+\frac{1}{2}}) \qquad \text{as } t \to 0, \tag{86}$$

where δ_{ij} is the Kronecker delta.

Proof. Notice that $\frac{1}{(2\pi at)^{m/2}} \exp\left(-\frac{\|x\|^2}{2at}\right)$ is the density function of a multivariate normal random variable $X \sim \mathcal{N}(0, atI)$. Equations (84) and (85) are simply the values of the coordinate-mean $\mathbb{E}[X_i] = 0$ and covariance $\mathrm{Cov}(X_i, X_j) = at\delta_{ij}$. Finally, we may write $X = \sqrt{at}Z$, where $Z \sim \mathcal{N}(0, I)$ is a standard multivariate normal (zero-mean, uncorrelated, unit variance) in m dimensions. We may write $\mathbb{E}[\|X\|^{2k+1}] = (at)^{k+\frac{1}{2}} \cdot \mathbb{E}[||Z||^{2k+1}]$, which confirms (86). □

Remark 16. *If instead of integrating over the entire domain $\mathbb{R}^m$, we integrate over a symmetric ball $B := B(R, 0)$ centered at zero, we recover the following augmented Gaussian identities:*

$$\frac{1}{(2\pi at)^{m/2}} \int_B x_i \exp\left(-\frac{\|x\|^2}{2at}\right) dx = 0, \tag{87}$$

$$\frac{1}{(2\pi at)^{m/2}} \int_B x_i x_j \exp\left(-\frac{\|x\|^2}{2at}\right) dx = \left(at + O(e^{-R^2/2at})\right)\delta_{ij} \qquad \text{as } t \to 0, \tag{88}$$

$$\frac{1}{(2\pi at)^{m/2}} \int_B \|x\|^{2k+1} \exp\left(-\frac{\|x\|^2}{2at}\right) dx = O(t^{k+\frac{1}{2}}) \qquad \text{as } t \to 0. \tag{89}$$

The restriction to the ball B leaves all odd-degree symmetries unchanged, and augments the even symmetries by a factor of the form $O(e^{-c/t})$, capturing the exponential decay on probability mass far away from the origin.

Lemma 4. *(Banach Mean Value Theorem) Consider Banach spaces $\mathcal{B}_1, \mathcal{B}_2$ and some open $\mathcal{U} \subseteq \mathcal{B}_1$, Then if $S : \mathcal{U} \to \mathcal{B}_2$ is Gateaux differentiable, then the mean value theorem holds in the sense that*

$$\|S(x) - S(y)\|_{\mathcal{B}_2} \leq \|x - y\|_{\mathcal{B}_1} \sup_{0 \leq t \leq 1} \|Df(tx + (1-t)y)\|$$

whenever the convex hull $[x, y]$ lies in $\mathcal{U}$.

Proof. The proof is a standard functional analysis argument but we recall it here for completeness. Let $\mathbb{L}$ be the one-dimensional subspace spanned by some fixed nonzero $u \in \mathcal{U}$. Consider $\varphi(cu) = c\|u\|$ as a continuous linear functional on $\mathbb{L}$ with norm 1. By the strong form of the Hahn-Banach, as stated in [3], for instance, we may extend this functional to the whole domain. Then note that $\varphi(u) = \|u\|$, so the result follows by considering $u = S(x) - S(y)$. □

Remark 17. *We recall that Fréchet differentiability in particular implies Gateaux differentiable, which will be sufficient for our purposes.*

Lemma 5. *(Banach Weak Law of Large Numbers) Let $\{X_j\}_{j \in \mathbb{N}}$ denote an independent identically distributed collection of random variables $X_j \in L^1(\Omega, B)$, where $\Omega := (\Omega, \Sigma, \mathbb{P})$ is a probability space and B is a Banach space. Let $S_n := \sum_{j=1}^n X_j$ denote the partial sum of the first n random variables. As $n \to \infty$, the normalized sequence $\frac{1}{n}S_n$ converges in probability to the mean $\mu := \mathbb{E}[X_j]$. That is for all $\epsilon > 0$,*

$$\lim_{n \to \infty} \mathbb{P}\left[\left|\frac{1}{n}S_n - \mu\right| > \epsilon\right] = 0.$$

Proof. See Pinelis [80] or Ledoux & Talagrand [64]. □

H.2 Key Lemmas for Theorem 1

Lemma 6. *(Taylor Series for Hilbert Signals) Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a Fréchet connection. For a given signal $S \in C^{n+1}(\mathcal{M}, \mathcal{E})$, the space of $(n+1)$-times continuously Fréchet-differentiable sections, and $p \in \mathcal{M}$, consider any $q \in \mathcal{M}$ in a geodesic ball of p and fix a length-minimizing curve $\gamma(t)$ from p to q. Let $S^*(t) := P_{\gamma(t)\to p} S(\gamma(t))$ denote the parallel transport of S from $\gamma(t)$ back to p along γ. As $t \to 0$, we have that*

$$S^*(t) = \left[\sum_{j=0}^{n} \frac{t^j}{j!} \big(\nabla_{\dot\gamma}^{(j)} S\big)(p)\right] + O(t^{n+1})$$

Proof. We first establish for a section $V(t)$ of $\mathcal{E}$ along γ, that:

$$\frac{d}{dt}\left(P_{\gamma(t)\to p} V(t)\right) = P_{\gamma(t)\to p}\left(\nabla_{\dot\gamma} V\right)(\gamma(t))$$

We compute directly from the definition:

$$\frac{d}{dt}\left(P_{\gamma(t)\to p} V(t)\right) = \lim_{h\to 0} \frac{P_{\gamma(t+h)\to p} V(\gamma(t+h)) - P_{\gamma(t)\to p} V(\gamma(t))}{h} \tag{90}$$

$$= \lim_{h\to 0} P_{\gamma(t)\to p}\left(\frac{P_{\gamma(t+h)\to\gamma(t)} V(\gamma(t+h)) - V(\gamma(t))}{h}\right) \tag{91}$$

$$= P_{\gamma(t)\to p}\left(\nabla_{\dot\gamma} V\right)(\gamma(t)) \tag{92}$$

where in the second line we used the composition law for parallel transport, $P^{-1}_{\gamma(t)\to p} \circ P_{\gamma(t+h)\to p} = P_{\gamma(t+h)\to\gamma(t)}$. which follows from uniqueness of the parallel transport ODE, together with the fact that $P_{\gamma(t)\to p}$ is a bounded linear operator and hence may be factored outside the limit.

Recall that $S^*(t) := P_{\gamma(t)\to p} S(\gamma(t))$ is a curve in the fixed Hilbert space $\mathcal{E}_p$. Iteratively applying the previous derivative computation to $V(t) = \big(\nabla_{\dot\gamma}^{(n)} S\big)(\gamma(t))$, where $(-)^{(n)}$ denotes n-fold composition, yields:

$$\frac{d^n}{dt^n} S^*(t) = P_{\gamma(t)\to p}\left(\nabla_{\dot\gamma}^{(n)} S\right)(\gamma(t))$$

Evaluating at $t = 0$, where $P_{p\to p} = \mathrm{id}$:

$$\left.\frac{d^n}{dt^n} S^*\right|_{t=0} = \left(\nabla_{\dot\gamma}^{(n)} S\right)_p .$$

Finally, applying Taylor's theorem for Banach spaces [25] yields the desired asymptotic statement. □

Lemma 7. *Let $\hat\Delta_{\mathcal{F}_n^t}$ and $\hat\Delta^t$ denote the point-cloud and functional Laplacian operators with bandwidth t. We have the concentration inequality:*

$$\mathbb{P}\left[\frac{1}{t(4\pi t)^{m/2}}\left|\hat\Delta_{\mathcal{F}_n^t} S(x) - \hat\Delta^t S(x)\right| > \epsilon\right] \le 2\exp\left(-\frac{t^2(4\pi t)^m \epsilon^2 n}{2K^2}\right)$$

for some $K > 0$ which depends only on the choice of section s. Consequently, we have the following limit in probability as $n \to \infty$:

$$\frac{1}{t(4\pi t)^{m/2}} \hat\Delta_{\mathcal{F}_n^t} S(x) \xrightarrow{\mathbb{P}} \hat\Delta^t S(x)\,.$$

Proof. The point cloud Laplacian $\Delta_{\mathcal{F}_n^t} S(x)$ may be viewed as the sample average of n iid Hilbert-space valued random variables:

$$X_i := \exp\left(-\frac{d_{\mathcal{M}}(x, x_i)^2}{4t}\right) S(x) - \exp\left(-\frac{d_{\mathcal{M}}(x, x_i)^2}{4t}\right) P_{x_i\to x} S(x_i)$$

Moreover, the functional approximation $\hat\Delta^t$ may be viewed as the expectation $\mathbb{E}\Delta_{\mathcal{F}_n^t}$ with respect to the uniform probability measure on $\mathcal{M}$. The bundle $\mathcal{E} \to \mathcal{M}$ has separable fibers, so the results of [80] apply. We may recover a Hoeffding inequality:

$$\mathbb{P}\left[\frac{1}{\delta}\left|\hat\Delta_{\mathcal{F}_n^t} S(x) - \mathbb{E}\hat\Delta_{\mathcal{F}_n^t} S(x)\right| > \epsilon\right] \le 2\exp\left(-\frac{(\delta\epsilon)^2 n}{2K^2}\right)$$

where K is the maximum norm of the section S over the compact manifold $\mathcal{M}$. Setting $\delta = t(4\pi t)^{m/2}$ and identifying $\mathbb{E}\hat{\Delta}_{\mathcal{F}_n^t} S(x) = \hat{\Delta}^t S(x)$ yields the desired concentration inequality. Convergence in probability follows immediately. □

Lemma 8. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle on a closed Riemannian manifold equipped with a Fréchet connection. Let $B \subseteq \mathcal{M}$ be open, and $p \in B$. Fix a section $S \in C^3(\mathcal{M}, \mathcal{E})$. For each $x \in \mathcal{M}$, let $F(x) := P_{x \to p} S(x)$ denote the parallel transport of $S(x)$ along the designated geodesic connecting x to p. For any real $a \in \mathbb{R}^{>0}$, the following asymptotic bound holds as $t \to 0$:*

$$\left| \int_{\mathcal{M}} e^{-\frac{\|y-p\|}{4t}} F(y)\, d\mu(y) - \int_{B} e^{-\frac{\|y-p\|}{4t}} F(y)\, d\mu(y) \right| = o(t^a).$$

Proof. We note this is a modified version of Lemma 4.1 of [14]. Let $d := \inf_{x \notin B} \|p - x\|$, $K := \sup_{x \in \mathcal{M}} \|S(x)\|$, and $M := \mu(\mathcal{M} \setminus B)$, where μ is the canonical measure with respect to the volume pseudo-form. Since p is compact and $\mathcal{M} \setminus B$ is closed, the infimum distance $d > 0$. Recalling that $P_{x \to p} : \mathcal{E}_x \to \mathcal{E}_p$ is unitary, and hence $\|F(x)\| = \|S(x)\|$ for all $x \in \mathcal{M}$, we may bound:

$$\begin{aligned}
\left| \int_{\mathcal{M}} e^{-\frac{\|y-p\|}{4t}} F(y)\, d\mu(y) - \int_{B} e^{-\frac{\|y-p\|}{4t}} F(y)\, d\mu(y) \right| &\leq \int_{\mathcal{M}\setminus B} \left\| e^{-\frac{\|y-p\|}{4t}} F(y) \right\| d\mu(y) \\
&\leq \int_{\mathcal{M}\setminus B} e^{-\frac{\|y-p\|}{4t}} \|S(y)\|\, d\mu(y) \\
&\leq MK \exp\left(-\frac{d}{4t}\right) \\
&= o(t^a).
\end{aligned}$$

□

Lemma 9. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a compatible Fréchet connection, with associated Laplacian Δ_∇, and functional approximation $\hat{\Delta}^t$ with bandwidth t . Fix a section $S \in C^3(\mathcal{M}, \mathcal{E})$. For any $x \in \mathcal{M}$ as the bandwidth $t \to 0$, we have pointwise convergence:*

$$\lim_{t \to 0} \frac{1}{t\,(4\pi t)^{\frac{m}{2}}} \hat{\Delta}^t S(x) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S(x)\,.$$

Proof. Let $\gamma_t := \frac{1}{t} \frac{1}{(4\pi t)^{m/2}}$, and consider the scaled functional approximation $\gamma_t \hat{\Delta}^t$ which acts on a section S at point p by:

$$\left(\gamma_t \hat{\Delta}^t S\right)(p) = \gamma_t \int_{\mathcal{M}} e^{-\frac{d_{\mathcal{M}}(p,x)^2}{4t}} (S(x) - P_{x \to p} S(p))\, d\mu(x)\,.$$

Let $B \subseteq \mathcal{M}$ denote a sufficiently small ball containing p. By Lemma 8,

$$\lim_{t \to 0} \left[\left(\gamma_t \hat{\Delta}^t S\right)(p) \right] = \lim_{t \to 0} \gamma_t \int_{B} e^{-\frac{d_{\mathcal{M}}(p,x)^2}{4t}} (S(p) - P_{p \to x} S(x))\, d\mu(x).$$

Parameterize B via geodesic coordinates such that $p = 0$. Let $F(x) := P_{x \to p} S(x)$. Let $\tilde{S} : \mathbb{R}^k \to \mathcal{H}$ and $\tilde{B} \subseteq \mathbb{R}^k$ denote the section S and ball B in coordinates. In these coordinates, we may write:

$$\lim_{t \to 0} \left[\left(\gamma_t \hat{\Delta}^t_{\mathcal{F}_n} S\right)(p) \right] = \lim_{t \to 0} \frac{1}{\mathrm{vol}(\mathcal{M})} \gamma_t \int_{\tilde{B}} e^{-\frac{d_{\mathcal{M}}(\exp_p(x),0)^2}{4t}} (\tilde{F}(0) - \tilde{F}(x)) \sqrt{\det(g_{ij})}\, dx\,.$$

In geodesic coordinates, since the closed manifold $\mathcal{M}$ has bounded Ricci curvature, the metric tensor has an asymptotic expansion given by (as in e.g. [36])

$$\det(g_{ij}) = 1 + O(\|x\|^2)\,.$$

This approximation and the identification $d_{\mathcal{M}}(\exp_p(x), p) = \|x\|$ in coordinates allows us to express:

$$\lim_{t \to 0} \left[\left(\gamma_t \hat{\Delta}^t_{\mathcal{F}_n} S\right)(p) \right] = \lim_{t \to 0} \frac{1}{\mathrm{vol}(\mathcal{M})} \gamma_t \int_{\tilde{B}} e^{-\frac{\|x\|^2}{4t}} (\tilde{F}(0) - \tilde{F}(x)) \left(1 + O(\|x\|^2)\right) dx\,.$$

Let $\mathbb{L}_t := \frac{1}{\mathrm{vol}(\mathcal{M})}\gamma_t \int_{\tilde{B}} \exp\left(-\frac{\|x\|^2}{4t}\right)(\tilde{F}(0)-\tilde{F}(x))\big(1+O(\|x\|^2)\big)\,dx$. This expression splits as $\mathbb{L}_t = A_t + B_t$, where:

$$A_t := \frac{1}{\mathrm{vol}(\mathcal{M})}\gamma_t \int_{\tilde{B}} \exp\left(-\frac{\|x\|^2}{4t}\right)(\tilde{F}(0)-\tilde{F}(x))\,dx\,,$$
$$B_t := \frac{1}{\mathrm{vol}(\mathcal{M})}\gamma_t \int_{\tilde{B}} \exp\left(-\frac{\|x\|^2}{4t}\right)(\tilde{F}(0)-\tilde{F}(x))\left[O(\|x\|^2)\right]\,dx\,.$$

We first analyze the limiting behavior of A_t as $t \to 0$. Within our geodesic coordinates centered at p, we may further work with a local synchronous frame of $\mathcal{E}$ along these coordinates $x = (x_1, \ldots, x_k)$ such that it is parallel along all radial geodesics. Consequently, within this frame, ordinary derivatives of $\tilde{F}$ coincide with covariant derivatives of S at the basepoint:

$$\partial_i \tilde{F}(p) = \nabla_{e_i} S(p)$$
$$\partial_i\partial_j \tilde{F}(p) = \nabla_{e_i}\nabla_{e_j} S(p) + \mathsf{curv}_{ij}\,,$$

where $\mathsf{curv}_{ij} := -\frac{1}{2}R^{\mathcal{E}}(e_i, e_j)S(p)$ is half the bundle curvature of $\mathcal{E}$ arising from the connection ∇. Hence by Lemma 6, the Taylor expansion of $\tilde{F}$ at p is

$$\tilde{F}(x) = \tilde{F}(p) + \sum_i x_i \nabla_{e_i} S(p) + \frac{1}{2}\sum_{i,j} x_i x_j\big(\nabla_{e_i}\nabla_{e_j} S(p) + \mathsf{curv}_{ij}\big) + O(\|x\|^3),$$

with $\mathsf{curv}_{ij} = -\mathsf{curv}_{ji}$.

Using the augmented Gaussian identities (87), (88), and (89), we may compute:

$$\begin{aligned} I_t &:= \frac{1}{(4\pi t)^{m/2}}\int_{\tilde{B}} (\tilde{F}(p)-\tilde{F}(x))\exp\left(-\frac{\|x\|^2}{4t}\right)dx \\ &= -\frac{1}{2}\sum_{i,j}\left[\left(\nabla_{e_i}\nabla_{e_j}S(p) + \mathsf{curv}_{ij}\right)\left(2t + O\left(e^{-c/t}\right)\right)\delta_{ij}\right] + O\left(t^{3/2}\right) \\ &= -t\sum_i \nabla_{e_i}\nabla_{e_i}S(p) - t\sum_i \mathsf{curv}_{ii} + O\left(t^{3/2}\right) \end{aligned}$$

where δ_{ij} is the Kronecker delta. Since curv_{ij} is antisymmetric, $\mathsf{curv}_{ii} = 0$, hence

$$I_t = -t\sum_i \nabla_{e_i}\nabla_{e_i} S(p) + O(t^{3/2}).$$

Inserting into the definition of A_t yields

$$\lim_{t\to 0} A_t = -\frac{1}{\mathrm{vol}(\mathcal{M})}\sum_i \nabla_{e_i}\nabla_{e_i} S(p).$$

Thus we recover the connection Laplacian by Lemma 1 .

To analyze the quantity B_t, we first observe that inside of $\tilde{B}$, the parallel transport map $P_{\exp_p(x)\to p}$ varies smoothly in x, with bounded derivatives. Since the section S is C^3, the mean value theorem (Lemma 4) ensures there is a $K > 0$ such that $\|F(x) - F(p)\| \le K\|x - p\|$ for all $x \in \tilde{B}$. Utilizing this Lipschitz bound and the augmented Gaussian identities, we may compute:

$$\begin{aligned} \|B_t\| &\le \frac{1}{\mathrm{vol}(\mathcal{M})}\gamma_t \int_{\tilde{B}} \exp\left(-\frac{\|x\|^2}{4t}\right)\|\tilde{F}(0)-\tilde{F}(x)\|\left[O(\|x\|^2)\right]\,dx \\ &\le \frac{K/\mathrm{vol}(\mathcal{M})}{t}\frac{1}{(4\pi t)^{m/2}}\int_{\tilde{B}} \|x\|^3 \exp\left(-\frac{\|x\|^2}{4t}\right) \\ &= \frac{1}{t}O\left(t^{3/2}\right) \\ &= O\left(\sqrt{t}\right)\,. \end{aligned}$$

Hence $B_t \to 0$ as $t \to 0$. Combining with the analysis of A_t yields:

$$\begin{aligned} \lim_{t\to 0}\left[\left(\gamma_t \hat{\Delta}^t_{\mathcal{F}_n} S\right)(p)\right] &= \lim_{t\to 0}[A_t + B_t] \\ &= \Delta_\nabla S(p) \end{aligned}$$

□

Lemma 10 (Variance asymptotics)**.** *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a compatible Fréchet connection, with associated Laplacian Δ_∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Consider a random sample of n-points with respect to the normalized volume pseudo-form, $\mathcal{X}_n = \{x_1, x_2, \cdots, x_n\} \subset \mathcal{M}$. For each bandwidth t, let $\gamma := (t(4\pi t)^{m/2})^{-1}$, where $m := \dim(\mathcal{M})$. Define an error term:*

$$\mathcal{R}_n(S) := \gamma \hat{\Delta}_{\mathcal{F}_n^t} S - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S \,.$$

There is a constant $C_{\mathrm{var}} > 0$, which depends on the section S, such that the following asymptotic estimate holds as $t \to 0^+$:

$$\mathbb{E}_{\mathcal{X}_n} \left[\| \mathcal{R}_n(S) - \mathbb{E}_{\mathcal{X}_n} \mathcal{R}_n(S) \|_{L^2}^2 \right] \leq \frac{C_{\mathrm{var}}}{n t^{2 + \frac{m}{2}}} \,.$$

Proof. Let Y_j be the random section given by

$$Y_j(x) := \gamma (S(x) - P_{x_j \to x} S(x_j)) \exp\left(- \frac{d_{\mathcal{M}}(x, x_j)^2}{4t} \right).$$

Note that Y_j is stochastic only through the sample point x_j, and that Y_i, Y_j are independent when $i \neq j$. It is straightforward to verify by Funbini's theorem that

$$\begin{aligned} \mathbb{E}_{\mathcal{X}_n} \left[\| \mathcal{R}_n(S) - \mathbb{E}_{\mathcal{X}_n} \mathcal{R}_n(S) \|_{L^2}^2 \right] &= \frac{1}{n} \left(\mathbb{E}_{x_1} \left[\| Y_1 \|_{L^2}^2 \right] - \| \mathbb{E}_{x_1} Y_1 \|_{L^2}^2 \right) \\ &\leq \frac{1}{n} \mathbb{E}_{x_1} \left[\| Y_1 \|_{L^2}^2 \right] \,. \end{aligned}$$

Set $K := \max_{x \in \mathcal{M}} \| S(x) \|_{\mathcal{E}_x}$. By Fubini, we may exchange the order of integration and find

$$\mathbb{E}_{x_1} \left[\| Y_1 \|_{L^2}^2 \right] = \int_{\mathcal{M}} \mathbb{E}_{x_1} \left[\| Y_j(x) \|_{\mathcal{E}_x}^2 \right] dx$$

We may compute:

$$\begin{aligned} \mathbb{E}_{x_1} \left[\| Y_1(x) \|_{\mathcal{E}_x}^2 \right] &= \frac{1}{\mathrm{vol}(\mathcal{M})} \int_{\mathcal{M}} \left\| \gamma (S(x) - P_{x_1 \to x} S(x_1)) \exp\left(- \frac{d_{\mathcal{M}}(x, x_1)^2}{4t} \right) \right\|_{\mathcal{E}_x}^2 dx_1 \\ &\leq \frac{4 \gamma^2 K^2}{\mathrm{vol}(\mathcal{M})} \int_{\mathcal{M}} \exp\left(- \frac{d_{\mathcal{M}}(x, x_1)^2}{2t} \right) dx_1 \,. \end{aligned}$$

By standard Gaussian identities, the remaining Gaussian integral is $O(t^{m/2})$ as $t \to 0^+$, with constant independent of x. Recalling the definition of γ in terms of t, we recover that

$$\mathbb{E}_{x_1} \left[\| Y_1(x) \|_{\mathcal{E}_x}^2 \right] = O(t^{-(2+m)}) \cdot O(t^{m/2}) = O(t^{-(2+m/2)}).$$

Since the constants are all independent of x, we may integrate over $\mathcal{M}$ and find that

$$\mathbb{E}_{x_1} \left[\| Y_1 \|_{L^2}^2 \right] = O(t^{-(2+m/2)})$$

as well. The result immediately follows. □

Lemma 11 (Bias asymptotics)**.** *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a compatible Fréchet connection, with associated Laplacian Δ_∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Consider a random sample of n-points with respect to the normalized volume pseudo-form, $\mathcal{X}_n = \{x_1, x_2, \cdots, x_n\} \subset \mathcal{M}$. For each bandwidth t, let $\gamma := (t(4\pi t)^{m/2})^{-1}$, where $m := \dim(\mathcal{M})$. Define an error term:*

$$\mathcal{R}_n(S) := \gamma \hat{\Delta}^t_{\mathcal{F}_n} S - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S$$

There is a constant $C_{\mathrm{bias}} > 0$, which depends on the section S, such that the following asymptotic estimate holds as $t \to 0^+$:

$$\left[\| \mathbb{E}_{\mathcal{X}_n} \mathcal{R}_n(S) \|_{L^2}^2 \right] \leq \left(C_{\mathrm{bias}} t \right)^2 .$$

Proof. This follows from essentially repeating the analysis in the proof of Lemma 9 using the fourth order Taylor expansion $F(y) = P_{y \to x} S(y)$ in terms of the covariant derivative. In particular, after accounting for the fourth-order Taylor remainder, we find that:

$$\mathbb{E}_{\mathcal{X}_n} [\gamma \hat{\Delta}_{\mathcal{F}_n^t}(x)] = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S(x) + O(t) \,.$$

The result follows. □

H.3 Proof of Theorem 1

H.3.1 Proof of Theorem 1A

Theorem. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a compatible Fréchet connection, with associated Laplacian Δ_∇. Fix a section $S \in C^3(\mathcal{M}, \mathcal{E})$. Consider a random sample of n-points with respect to the normalized volume form, $\mathcal{X}_n = \{x_1, x_2, \cdots, x_n\} \subset \mathcal{M}$. Let $\mathcal{F}^t_{\mathcal{X}_n}$ be the associated Hilbert cellular sheaf with bandwidth t and associated Point cloud Laplacian $\hat{\Delta}_{\mathcal{F}^t_n}$. Then, we have that in probability, for any $x \in \mathcal{M}$,*

$$\lim_{n\to\infty} \frac{1}{t_n \left(4\pi t_n\right)^{\frac{m}{2}}} \hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S(x)$$

with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$.

Proof. Let $\gamma_n := \frac{1}{t_n} \frac{1}{(4\pi t_n)^{m/2}}$, and consider the scaled functional approximation $\gamma_n \hat{\Delta}^{t_n}$ which acts on a section S at point p by:

$$\left(\gamma_n \hat{\Delta}^{t_n} S\right)(p) = \gamma_n \int_{\mathcal{M}} e^{-\frac{d_{\mathcal{M}}(p,x)^2}{4t_n}} \left(P_{x\to p} S(x) - S(p)\right) d\mu(x)\,.$$

We may bound:

$$\begin{aligned}\mathbb{P}\left[\left\|\frac{1}{\gamma_n}\hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S(x)\right\| > 2\epsilon\right] \leq \quad & \mathbb{P}\left[\frac{1}{\gamma_n}\left\|\hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) - \hat{\Delta}^{t_n} S(x)\right\| > \epsilon\right] \\ & + \mathbb{P}\left[\left\|\frac{1}{\gamma_n}\hat{\Delta}^{t_n} S(x) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S(x)\right\| > \epsilon\right]\end{aligned}$$

as $n \to \infty$ we have $t_n \to 0$. Hence the second quantity on the right hand side goes to zero by Lemma 9. On the other hand, the first quantity on the right hand side can be bound by the concentration inequality of Lemma 7, yielding:

$$\mathbb{P}\left[\frac{1}{\gamma_n}\left\|\hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) - \hat{\Delta}^{t_n} S(x)\right\| > \epsilon\right] \leq 2\exp\left(\frac{t_n^{2+m} n (4\pi t)^m \epsilon^2}{2K^2}\right),$$

where K is a constant depending on the section S. Since $nt_n^{2+m} \to \infty$ as $n \to \infty$, the concentration upper bound goes to zero as $n \to \infty$ as well. This completes the proof of the of the main theorem. □

H.3.2 Proof of theorem 1B

Theorem. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle equipped with a compatible Fréchet connection, with associated Laplacian Δ_∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Consider a random sample of n-points with respect to the normalized volume form, $\mathcal{X}_n = \{x_1, x_2, \cdots, x_n\} \subset \mathcal{M}$. Let $\mathcal{F}^t_{\mathcal{X}_n}$ be the associated Hilbert cellular sheaf with bandwidth t and associated Point cloud Laplacian $\hat{\Delta}_{\mathcal{F}^t_n}$. Then, we have the following convergence in expectation:*

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X}}\left[\left\|\frac{1}{t_n\left(4\pi t_n\right)^{\frac{m}{2}}}\hat{\Delta}_{\mathcal{F}^{t_n}_n} S(x) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S(x)\right\|^2_{L^2}\right] = 0$$

with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$.

Proof. Let $\gamma_n := (t_n(4\pi t_n)^{m/2})^{-1}$. Define an error term:

$$\mathcal{R}_n(S) := \gamma \hat{\Delta}_{\mathcal{F}^t_n} S - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S\,.$$

We may decompose the error into a bias and variance term as:

$$\mathbb{E}_{\mathcal{X}_n}\left[\|\mathcal{R}_n\|^2_{L^2}\right] = \mathbb{E}_{\mathcal{X}_n}\left[\|\mathcal{R}_n(S) - \mathbb{E}_{\mathcal{X}_n}\mathcal{R}_n(S)\|^2_{L^2}\right] + \|\mathbb{E}_{\mathcal{X}_n}\mathcal{R}_n(S)\|^2_{L^2}\,.$$

By the asymptotic results of Lemmas 10 and 11, there are positive constants C_{var} and C_{bias} such that

$$\mathbb{E}_{\mathcal{X}_n}\left[\|\mathcal{R}_n\|^2_{L^2}\right] \leq \frac{C_{\mathrm{var}}}{nt_n^{2+\frac{m}{2}}} + C^2_{\mathrm{bias}} t_n^2$$

Since $t_n = n^{-1/(m+2+\alpha)}$, we have $t_n^2 \to 0$ and $nt_n^{2+\frac{m}{2}} \to \infty$ as $n \to \infty$. Hence both terms vanish in limit, proving the desired convergence. □

H.4 Key Lemmas for Theorem 2

Definition 25. *Let $\mathcal{E}$ be a smooth Hilbert bundle over a manifold $\mathcal{M}$. A **finite rank approximating sequence** for $\mathcal{E}$ is a sequence of smooth sub-bundles $\{\mathcal{E}_d\}_{d\geq 1}$ with the following properties:*

1. *For each d, $\mathcal{E}_d$ has finite rank;*
2. *For each d, the bundle $\mathcal{E}_d$ is a sub-bundle of $\mathcal{E}_{d+1}$;*
3. *For each $x \in \mathcal{M}$, we have that $\mathrm{cl}\left(\mathrm{span}\left(\bigcup_d (\mathcal{E}_d)_x\right)\right) = \mathcal{E}_x$.*

Lemma 12. *Let $\mathcal{E}$ be a smooth Hilbert bundle with infinite-dimensional fibers over a compact manifold $\mathcal{M}$. A finite rank approximating sequence $\{\mathcal{E}_d\}_d$ exists.*

Proof. By Kuiper's Theorem [61], the unitary group of the typical fiber $\mathcal{H}$ is contractible, implying that there exists an isomorphism of $\mathcal{E}$ with bundle $\mathcal{M} \times \mathcal{H}$ at the level of purely topological bundles. Now note that every Hilbert bundle $\mathcal{M} \times \mathcal{H}$ admits a finite rank approximating sequence, by considering a Hilbert space basis $\{e_1, e_2, \ldots\}$ for $\mathcal{H}$, and defining $\mathcal{H}_n := \mathrm{span}(e_1, \ldots, e_n)$. The sequence $\{\mathcal{M} \times \mathcal{H}_d\}_d$ can then be seen to be a finite rank approximating sequence. Furthermore, because the base space is a finite-dimensional manifold, this topological trivialization can be upgraded to a smooth global trivialization [75]. Let $\Phi : \mathcal{E} \to \mathcal{M} \times \mathcal{H}$ be such a smooth isomorphism. Thus, the finite rank approximating sequence $\{\mathcal{M} \times \mathcal{H}_d\}_d$ pulls back to a finite rank approximating sequence on $\mathcal{E}$ by $\mathcal{E}_d := \Phi^{-1}(\mathcal{M} \times \mathcal{H}_d)$, as desired. □

Lemma 13. *Let $(\mathcal{E}, \mathcal{M}, \nabla)$ be a smooth infinite-dimensional Hilbert bundle over a compact manifold $\mathcal{M}$ equipped with a compatible connection ∇. Let $\{\mathcal{E}_d\}_{d\geq 1}$ be a finite rank approximating sequence for $\mathcal{E}$. The data of ∇ induces a compatible connection $\nabla_d := \Pi_d \nabla$ on $\mathcal{E}_d$, where $\Pi_d : \mathcal{E} \to \mathcal{E}_d$ denotes the fiber-wise orthogonal projection onto $\mathcal{E}_d$.*

Proof. This follows immediately from the fact that ∇ is compatible and orthogonal projections are self-adjoint. □

Remark 18. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be an infinite-dimensional Hilbert bundle equipped with a compatible connection. By the previous lemmas, we may always find a finite rank approximating sequence $(\mathcal{E}_d, \mathcal{M}, \nabla_d)$, each with compatible connection. These compatible connections induce connection Laplacians Δ_{∇_d} on each sub-bundle. Moreover, Theorem 1B applies to each Laplacian Δ_{∇_d}.*

We restate and prove Proposition 1.

Proposition. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a Hilbert bundle, with strictly infinite-dimensional generic Hilbert-space fiber $\mathcal{H}$. Fix an orthogonal basis $\mathcal{B} = \{e_1, e_2, ...\}$ of $\mathcal{H}$ and let $\mathcal{H}_d := \mathrm{span}(e_1, e_2, ..., e_d)$. Then there exists a smooth map of bundles*

$$\Pi_d : \mathcal{E} \to \mathcal{E}_d \tag{93}$$

where $\mathcal{E}_d$ is a d-dimensional vector bundle with generic fiber $\mathcal{H}_d$ and at each $x \in \mathcal{M}$, $\Pi_d|_{\mathcal{E}_x} : \mathcal{E}_x \to \mathcal{E}_{d,x}$ recovers the usual orthogonal projection map.

Proof. One we fix a basis $\mathcal{B}$, the conclusion follows from the proof of Lemma 12 and an application of Lemma 13. □

Definition 26. *Let $(\mathcal{M}, \mathcal{E}, \nabla)$ be a smooth Hilbert bundle over a closed manifold $\mathcal{M}$ of dimension m equipped with a compatible connection ∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Let $\{\mathcal{E}_d\}_d$ be a finite rank approximating sequence for $\mathcal{E}$ with induced connections ∇_d, connection Laplacians Δ_{∇_d}, and bandwidth t point cloud Laplacians $\hat{\Delta}^t_{n,d}$ associated to an iid sampling $\mathcal{X} = \{x_1, x_2, \ldots\}$. Let $\Pi_d : \mathcal{E} \to \mathcal{E}_d$ denote the fiber-wise orthogonal projection map onto $\mathcal{E}_d$. The **discretization error** $\mathtt{D}(n, d)$ and the **continuous geometry error** $\mathtt{E}(d)$ are the quantities:*

$$\mathsf{D}(n,d) := \left\| \gamma_n \hat{\Delta}^{t_n}_{n,d}(\Pi_d S) - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\nabla_d}(\Pi_d S) \right\|^2_{L^2}$$

$$\mathsf{E}(d) := \left\| \frac{1}{\mathrm{vol}(\mathcal{M})} \left(\Delta_{\nabla_d}(\Pi_d S) - \Delta_\nabla S\right) \right\|^2_{L^2}.$$

Remark 19. *The discretization error $\mathsf{D}(n, d)$ captures the error introduced by approximating the connection Laplacian Δ_{∇_d} by a point-cloud Laplacian on n points. The continuous geometry error $\mathsf{E}(d)$ is a deterministic quantity that captures the error introduced by moving to the sub-bundle $\mathcal{E}_d$.*

Lemma 14. *The continuous geometry error* $\mathsf{E}(d)$ *converges to zero as* $d \to \infty$.

Proof. Without loss of generality, by pushing through the global trivialization of Kuiper's theorem [61], we may assume without loss of generality that $\mathcal{E} = \mathcal{H} \times \mathcal{M}$ as topological bundles. First, note that the orthogonal projection Π_d commutes with the Fréchet derivative D on sections, in the sense that:

$$D(\Pi_d S) = \Pi_d(DS).$$

By Proposition 2, for any vector field X, we may write:

$$\nabla_X S = DS(X) + A(X)S$$

where $A(X)$ is a globally bounded linear operator acting on the fiber $\mathcal{H}$. We may now compute:

$$\begin{aligned}(\nabla_d)_X(\Pi_d S) &= \Pi_d \nabla_X(\Pi_d S) \\ &= \Pi_d \left(D\Pi_d S(X) + A(X)\Pi_d S\right) \\ &= \Pi_d DS(X) + \Pi_d A(X)\Pi_d S)\end{aligned}$$

Since $\Pi_d \to \mathrm{id}$ in the strong operator topology and $DS(X) \in \mathcal{H}$, we have that $\Pi_d DS(X) \to DS(X)$ as $d \to \infty$. Similarly since $A(X) : \mathcal{H} \to \mathcal{H}$ is bounded, we find that for each $x \in \mathcal{M}$, we have $\Pi_d A(X)\Pi_d S(x) \to A(X)S(x)$. Therefore $\nabla_d(\Pi_d S)(x) \to \nabla S(x)$. By a similar argument, we may conclude that for a pair of vector fields X, Y, that $(\nabla_d)_X(\nabla_d)_Y(\Pi_d S)(x) \to \nabla_X \nabla_Y S(x)$. Hence using the coordinate form of the connection Laplacian of Lemma 1, we recover that:

$$\lim_{d\to\infty} [\Delta_{\nabla_d} S(x)] = \Delta_\nabla S(x)$$

for all $x \in \mathcal{M}$.

We now upgrade to this statement to L^2 convergence by the dominated convergence theorem. If there is a global bound K such that $\|\Delta_{\nabla_d}(\Pi_d S)(x) - \Delta_\nabla S(x)\|_{\mathcal{H}} \leq K$ for all $x \in \mathcal{M}$, we may apply the dominated convergence theorem and conclude that $E(d) \to 0$. To find such a K, we first observe that $\Delta_\nabla S$ is a continuous section, and hence $\|\Delta_\nabla S(x)\|_{\mathcal{H}} \leq K_1$ is bounded on the compact manifold $\mathcal{M}$. Next, for a pair of vector fields X, Y, we may use the fact that $\|\Pi_d\|_{\mathrm{op}} = 1$ and the representation $\nabla = D + A$ to find a d-independent bound on $\left\|\left((\nabla_d)_X(\nabla_d)_Y(\Pi_d S)\right)(x)\right\|_{\mathcal{H}}$. The local coordinate form the connection Laplacian Δ_{∇_d} again allows us to conclude that there is a bound $\|\Delta_{\nabla_d}(\Pi_d S)(x)\|_{\mathcal{H}} \leq K_2$ for all $x \in \mathcal{M}$ and $d \geq 1$. The triangle inequality finally allows us to bound:

$$\|\Delta_{\nabla_d}(\Pi_d S)(x) - \Delta_\nabla S(x)\|_{\mathcal{H}} \leq K_1 + K_2\,.$$

This completes the proof. □

H.5 Proof of Theorem 2

Theorem. *Let* $(\mathcal{E}, \mathcal{M}, \nabla)$ *be an infinite-dimensional Hilbert bundle over a closed manifold* $\mathcal{M}$ *of dimension* m *equipped with a compatible connection* ∇*. Fix a section* $S \in C^4(\mathcal{M}, \mathcal{E})$*. Let* $\{\mathcal{E}_d\}_d$ *be a finite rank approximating sequence for* $\mathcal{E}$ *with induced connections* ∇_d*, connection Laplacians* Δ_{∇_d}*, and bandwidth* t *point cloud Laplacians* $\hat{\Delta}_{\mathcal{F}^t_{n,d}}$ *associated to an iid sampling* $\mathcal{X} = \{x_1, x_2, \ldots\}$*. Let* $\Pi_d : \mathcal{E} \to \mathcal{E}_d$ *denote the fiber-wise orthogonal projection map onto* $\mathcal{E}_d$*. There exists a deterministic increasing sequence* d_n*, depending on the section* S*, such that*

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X}} \left[\left\| \frac{1}{t_n(4\pi t_n)^{m/2}} \hat{\Delta}_{\mathcal{F}^{t_n}_{n,d_n}} (\Pi_{d_n} S) - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S \right\|^2_{L^2} \right] = 0$$

with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$*,* $\alpha > 0$*.*

Proof. Let $\gamma_n := \frac{1}{t_n(4\pi t_n)^{m/2}}$. We may easily bound the expected global error in terms of the continuous geometry error and the expected discretization error:

$$\mathbb{E}_{\mathcal{X}} \left[\left\| \gamma_n \hat{\Delta}^{t_n}_{\mathcal{F}_{n,d_n}} (\Pi_{d_n} S) - \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_\nabla S \right\|^2_{L^2} \right] \leq 2\mathbb{E}_{\mathcal{X}_n}[\mathsf{D}(n,d)] + 2\mathsf{E}(d)\,.$$

By applying Theorem 1 to the bundle $\mathcal{E}_d$, the expected discretization error $\mathbb{E}_{\mathcal{X}_n}[\mathsf{D}(n,d)] \to 0$ as $n \to \infty$. On the other hand, Lemma 14 ensures that $\mathsf{E}(d) \to 0$ as $d \to \infty$.

To construct the diagonal sequence, first choose an increasing sequence p_i such that $\mathsf{E}(d) < \frac{1}{i}$ for all $d \geq p_i$. Next, choose an increasing sequence N_i such that $\mathbb{E}_{\mathcal{X}_n}[\mathsf{D}(n, p_i)] < \frac{1}{i}$ for each $n \geq N_i$. For each $n \geq 1$, set $\phi(n) := \max\{i \mid N_i \leq n\}$. Observe that both $\mathsf{E}(p_{\phi(n)}) < \frac{1}{\phi(n)}$ and $\mathbb{E}_{\mathcal{X}_n}[\mathsf{D}(n, p_{\phi(n)})] < \frac{1}{\phi(n)}$. Therefore setting $d_n := p_{\phi(n)}$ yields a diagonal sequence with the property that

$$\mathbb{E}_{\mathcal{X}}\left[\left\|\gamma_n \hat{\Delta}_{\mathcal{F}^{t_n}_{n,d_n}}(\Pi_{d_n} S) - \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla S\right\|^2_{L^2}\right] < \frac{4}{\phi(n)}\,,$$

which tends to zero since $\phi(n) \to \infty$. This diagonal subsequence is deterministic as both the continuous geometry error and the *expected* discretization error are deterministic. □

H.6 Key Lemmas for Corollary 2.1

Lemma 15. *Let $(\mathcal{E}, \mathcal{M}, \nabla)$ be a smooth infinite-dimensional Hilbert bundle over a closed manifold $\mathcal{M}$ of dimension m equipped with a compatible connection ∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Let $\{\mathcal{E}_d\}_d$ be a finite rank approximating sequence for $\mathcal{E}$ with induced connections ∇_d, connection Laplacians Δ_{∇_d}, and bandwidth t point cloud Laplacians $\hat{\Delta}_{\mathcal{F}^t_{n,d}}$ associated to an iid sampling $\mathcal{X} = \{x_1, x_2, \ldots\}$. Let $\Pi_d : \mathcal{E} \to \mathcal{E}_d$ denote the fiber-wise orthogonal projection map onto $\mathcal{E}_d$. Let $\{d_n\}$ be the diagonal sequence induced by Theorem 2. Let $\tilde{\Delta}_n := \frac{1}{t_n(4\pi t_n)^{m/2}}\hat{\Delta}_{\mathcal{F}^{t_n}_{n,d_n}}\Pi_{d_n}$ with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$. Similarly, let $\tilde{\Delta} := \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla$. Finally, let $g : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function.*

Under the Borel functional calculus (Appendix G.4), we have MSE convergence:

$$\mathbb{E}\left[\left\|g(\tilde{\Delta}_n)S - g(\tilde{\Delta})S\right\|^2_{L^2}\right] \to 0\,.$$

Proof. Observe that $\tilde{\Delta}$ and each $\tilde{\Delta}_n$ are self-adjoint unbounded operators on $L^2(\mathcal{E}; \mathcal{M})$. We begin by showing there is a common core $\mathcal{D} \subseteq L^2(\mathcal{E}; \mathcal{M})$ and a subsequence n_i for which $\tilde{\Delta}_{n_i} S \to \tilde{\Delta} S$ in $L^2(\mathcal{E}; \mathcal{M})$.

Take $\mathcal{D}$ to be any countable dense subset of $C^\infty(\mathcal{M}, \mathcal{E})$. Since $\mathcal{E}$ has separable fibers, such a countable dense subset necessarily exists. Moreover, $\tilde{\Delta}$ and $\tilde{\Delta}_n$ are all defined on $\mathcal{D}$. This $\mathcal{D}$ shall be our common core.

Treat $\tilde{\Delta}_n S$ as an $L^2(\mathcal{M}, \mathcal{E})$-valued random variable. For each $S \in \mathcal{D}$, Theorem 2 ensures that $\tilde{\Delta}_n S \to \tilde{\Delta} S$ in mean square error. It immediately follows that $\tilde{\Delta}_n S \xrightarrow{\mathbb{P}_\mathcal{X}} \tilde{\Delta} S$ in probability with respect to the measure from which the sampling $\mathcal{X}$ is drawn.

Enumerate $\mathcal{D} = \{S_1, S_2, \ldots\}$. Since convergence in probability implies almost sure convergence on a subsequence, we may inductively construct a doubly-indexed sequence of indices N^a_b such that the following properties hold:

1. For each a, the sequence $\{N^{a+1}_b\}_b$ is a subsequence of $\{N^a_b\}_b$;
2. Along the sequence $\{N^a_b\}_b$, we have almost sure convergence $\tilde{\Delta}_{N^a_b} S_a \xrightarrow{\text{a.s.}} \tilde{\Delta} S_a$ as $b \to \infty$.

Take the diagonal sequence $n_i := N^i_i$. Along n_i, we have that $\tilde{\Delta}_{n_i} S \xrightarrow{\text{a.s.}} \tilde{\Delta} S$ as $i \to \infty$ for all $S \in \mathcal{D}$. Now applying Theorems VIII.25(a) and VIII.20(b) of [82], we may conclude that:

$$\|g(\tilde{\Delta}_{n_i})S - g(\tilde{\Delta})S\|_{L^2} \xrightarrow{\text{a.s.}} 0$$

almost surely for each S.

Notice that the previous argument can not only be applied to the sequence $n = \{1, 2, 3, ...\}$, but also along any subsequence of $\{n\}_n$. Since any subsequence of $\{n\}_n$ therefore has an almost surely convergent sub-subsequence, we may conclude that for the original sequence n, for each section S (not necessarily in $\mathcal{D}$), we have convergence in probability:

$$\|g(\tilde{\Delta}_n)S - g(\tilde{\Delta})S\|_{L^2} \xrightarrow{\mathbb{P}_\mathcal{X}} 0$$

in probability with respect to the sampling $\mathcal{X}$ as $n \to \infty$.

Finally, by the spectral calculus, since $g : \mathbb{R} \to \mathbb{R}$ is bounded by some B, we may bound $\|g(\tilde{\Delta}_n)S\|_{L^2} \leq B\|S\|_{L^2}$. Similarly, $|g(\tilde{\Delta})S\|_{L^2} \leq B\|S\|_{L^2}$. It follows that for each n,

$$\|g(\tilde{\Delta}_n)S - g(\tilde{\Delta})S\|_{L^2}^2 \leq 4B^2$$

for all n. Hence the dominated convergence theorem admits an MSE upgrade to the desired conclusion:

$$\mathbb{E}\left[\left\|g(\tilde{\Delta}_n)S - g(\tilde{\Delta})S\right\|_{L^2}^2\right] \to 0\,.$$

□

H.7 Proof of Corollary 2.1

Corollary. *Under the hypotheses of Theorem 2, let $\{d_n\}_n$ be the constructed deterministic diagonal sequence, and $L \in \mathbb{N}$. Let σ be a fiber-wise nonlinearity that is C_σ-Lipschitz in the corresponding fiber norms. For bounded continuous filters $g^0, \ldots, g^{L-1} \in \mathcal{W}$, consider the continuous and sampled architectures:*

$$S^{\ell+1} := \sigma\left(g^\ell\left(\tilde{\Delta}\right)S^\ell\right)$$
$$S_n^{\ell+1} := \sigma\left(g^\ell\left(\tilde{\Delta}_n\right)S_n^\ell\right)\,,$$

with initializations $S^0 := S$ and $S_n^0 := \Pi_{d_n}S$, where $\tilde{\Delta} := \frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_\nabla$ and $\tilde{\Delta}_n := \frac{1}{t_n(4\pi t_n)^{m/2}}\hat{\Delta}_{\mathcal{F}_{n,d_n}^{t_n}}$.
Then, the output of the discrete architecture converges in mean square to the output of the continuous architecture:

$$\lim_{n\to\infty}\mathbb{E}\left[\left\|S_n^L - S^L\right\|_{L^2(\mathcal{M};\mathcal{E})}^2\right] = 0\,.$$

Proof. We proceed by induction on the layers. Let $e_{n,\ell}$ and $\delta_{n,\ell}$ denote the signal error and spectral filter error at layer ℓ respectively:

$$e_{n,\ell} := \left\|S_n^\ell - S^\ell\right\|_{L^2}$$
$$\delta_{n,\ell} := \left\|g_\ell(\tilde{\Delta}_n)(\Pi_{d_n}S^\ell) - g_\ell(\tilde{\Delta})S^\ell\right\|_{L^2}\,.$$

Let $M_\ell := \sup_{x\in\mathbb{R}}\|g_\ell(x)\|$. By the Lipschitz continuity of the nonlinearity σ, the triangle inequality yields the pathwise recursive bound:

$$e_{n,\ell+1} \leq C_\sigma(M_\ell e_{n,\ell} + \delta_{n,\ell})\,.$$

Iterating this inequality over L layers expands to:

$$e_{n,L} \leq \left(\prod_{r=0}^{L-1} C_\sigma M_r\right)e_{n,0} + \sum_{q=0}^{L-1}\left(\prod_{r=q+1}^{L-1} C_\sigma M_r\right)C_\sigma\delta_{n,q}\,.$$

By Theorem 2, the initialized signal error $e_{n,0} \to 0$ in mean square. Furthermore, by Lemma 15, $\delta_{n,q} \to 0$ in mean square as well for each q.
Taking the expectation of the squared recursive bound and applying the Cauchy-Schwarz inequality to the finite sum isolates the individual mean square limits. Hence, as $n \to 0$, we have that the total error satisfies:

$$\lim_{n\to\infty}\mathbb{E}[e_{n,L}^2] = 0\,.$$

In particular, we have that as sampling density goes to infinity, in MSE ,

$$\Omega(\mathcal{F}_{n,d_n}, \hat{\Delta}_{\mathcal{F}_{n,d_n}}, \mathcal{W}, \sigma) \to \Omega(\mathcal{E}, \Delta_\nabla, \mathcal{W}, \sigma)\,.$$

□

H.8 Proof of Corollary 2.2

Corollary. *Let $(\mathcal{E}, \mathcal{M}, \nabla)$ be a smooth Hilbert bundle over a closed manifold $\mathcal{M}$ of dimension m equipped with a compatible connection ∇. Fix a section $S \in C^4(\mathcal{M}, \mathcal{E})$. Let $\{\mathcal{E}_d\}_d$ be a finite rank approximating sequence for $\mathcal{E}$ with induced connections ∇_d, and connection Laplacians Δ_{∇_d}. Let $\Pi_d : \mathcal{E} \to \mathcal{E}_d$ denote the fiber-wise orthogonal projection map onto $\mathcal{E}_d$. Let $\mathcal{X} = \{x_1, x_2, \ldots\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots\}$ be a pair of independent iid samplings of points on the manifold. Denote the bandwidth t point cloud Laplacians associated to these distinct samplings by $\hat{\Delta}_{\mathcal{F}^t_{\mathcal{X}_n,d}}$ and $\hat{\Delta}_{\mathcal{F}^t_{\mathcal{Y}_n,d}}$ respectively. Let $\{d_n\}$ be a diagonal sequence such that the conclusion of Theorem 2 holds for both samplings. Let $\tilde{\Delta}^{\mathcal{X}}_n := \frac{1}{t_n(4\pi t_n)^{m/2}} \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{X}_{n,d_n}}} \Pi_{d_n}$ with bandwidth $t_n = n^{-\frac{1}{m+2+\alpha}}$, $\alpha > 0$, and similar for $\tilde{\Delta}^{\mathcal{Y}}_n$.*

Let $L \in \mathbb{N}$ be a network depth, and σ be a fiber-wise nonlinearity that is C_σ-Lipschitz in the corresponding fiber norms. For bounded continuous filters $g_0, \ldots, g_{L-1} \in \mathcal{W}$, consider the continuous and sampled architectures:

$$\begin{aligned} S^{\ell+1} &:= \sigma\left(g_\ell\left(\tilde{\Delta}\right) S^\ell\right) \\ S^{\mathcal{X},\ell+1}_n &:= \sigma\left(g_\ell\left(\tilde{\Delta}^{\mathcal{X}}_n\right) S^{\mathcal{X},\ell}_n\right) \\ S^{\mathcal{Y},\ell+1}_n &:= \sigma\left(g_\ell\left(\tilde{\Delta}^{\mathcal{Y}}_n\right) S^{\mathcal{Y},\ell}_n\right) , \end{aligned}$$

with initializations $S^0 := S$ and $S^{\mathcal{X},0}_n, S^{\mathcal{Y},0}_n := \Pi_{d_n} S$. Under these hypotheses and notation, one may obtain a MSE convergence result:

$$\lim_{n\to\infty} \mathbb{E}\left[\left\|S^{\mathcal{X},L}_n - S^{\mathcal{Y},L}_n\right\|^2_{L^2}\right] = 0\,.$$

Further, one may derive a quantitative bound for the L^2 disagreement $\|S^{\mathcal{X},L}_n - S^{\mathcal{Y},L}_n\|_{L^2}$ in terms of sample-indpendent quantities.

Proof. One may bound:

$$\left\|S^{\mathcal{X},L}_n - S^{\mathcal{Y},L}_n\right\|^2_{L^2} \leq\ 2\left\|S^{\mathcal{X},L}_n - S^L\right\|^2_{L^2} + 2\left\|S^{\mathcal{Y},L}_n - S^L\right\|^2_{L^2}$$

Applying Corollary 2.1 to each sampling separately yields the MSE convergence. In particular, we have that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{X},\mathcal{Y}}\left[\left\|\Omega(\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{X}_n,d_n}}, \mathcal{W}, \sigma) - \Omega(\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}, \hat{\Delta}_{\mathcal{F}^{t_n}_{\mathcal{Y}_n,d_n}}, \mathcal{W}, \sigma)\right\|^2_{L^2}\right] = 0\,.$$

To derive a quantitative bound, introduce the per-sampling signal error and spectral filter error at level ℓ by:

$$\begin{aligned} e^{(-)}_{n,\ell} &:= \left\|S^{(-),\ell}_n - S^\ell\right\|_{L^2} \\ \delta^{(-)}_{n,\ell} &:= \left\|h_\ell(\tilde{\Delta}^{(-)}_n)(\Pi_{d_n} S^\ell) - h_\ell(\tilde{\Delta}) S^\ell\right\|_{L^2} . \end{aligned}$$

Apply the triangle inequality and the layer-wise recursive bounds of the proof of Corollary 2.1 to establish:

$$\left\|S^{\mathcal{X},L}_n - S^{\mathcal{Y},L}_n\right\|_{L^2} \leq \left(\prod_{r=0}^{L-1} C_\sigma M_r\right)(e^{\mathcal{X}}_{n,0} + e^{\mathcal{Y}}_{n,0}) + \sum_{q=0}^{L-1}\left(\prod_{r=q+1}^{L-1} C_\sigma M_r\right) C_\sigma(\delta^{\mathcal{X}}_{n,q} + \delta^{\mathcal{Y}}_{n,q})\,.$$

The level-zero signal error is sampling independent, and bounded above by $\|S\|_{L^2}$. On the other hand, we may further apply the Borel functional calculus to bound each spectral filter error by:

$$\delta^{(-)}_{n,q} \leq 2M_q \|S^q\|_{L^2}\,.$$

We hence conclude a sample-independent bound:

$$\left\|S^{\mathcal{X},L}_n - S^{\mathcal{Y},L}_n\right\|_{L^2} \leq 2\left(\prod_{r=0}^{L-1} C_\sigma M_r\right)\|S\|_{L^2} + 2\sum_{q=0}^{L-1}\left(\prod_{r=q+1}^{L-1} C_\sigma M_r\right) C_\sigma M \|S^q\|_{L^2}\,.$$

□

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction state four contributions: (i) convolution over Hilbert bundles via the connection Laplacian, (ii) HilbNets architecture, (iii) discretization via Hilbert cellular sheaves with convergence guarantees generalizing Belkin & Niyogi, and (iv) experimental validation. All four are delivered. The scope is accurately represented; the authors note the framework is general while experiments focus on synthetic and traffic domains.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: An exhaustive limitations paragraph is presented in Section 7.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

Answer: [Yes]

Justification: All theorems, propositions, corollaries, and lemmas are numbered and cross-referenced. Assumptions (closed Riemannian manifold, Fréchet connection, C^3/C^4 regularity, i.i.d. uniform sampling, bandwidth schedule) are stated in theorem statements. Complete proofs are provided in the appendix with supporting lemmas. The heat kernel existence proof is a sketch referencing Berline–Getzler–Vergne, but the convergence theorems do not depend on it directly.

Guidelines:

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

4. **Experimental result reproducibility**

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

Answer: [Yes]

Justification: The appendix provides full experimental details: dataset descriptions, graph construction procedures, model architectures, all hyperparameters (optimizer, learning rate, batch size, polynomial filter order, number of Householder reflections, regularization weight, early stopping), number of seeds, and loss functions. The synthetic bundle is specified in closed form. Standard public datasets (METR-LA, PEMS-BAY) are used with standard splits.

Guidelines:

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

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [Yes]

   Justification: Code is publicly available at `https://github.com/clabat9/hilbnet`. The datasets used (METR-LA, PEMS-BAY) are publicly available and properly cited. The paper provides sufficient algorithmic and architectural detail for reimplementation.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: Core experimental setup is summarized in Section 6. Full details including data splits (standard 70/10/20 chronological partition), optimizer (Adam), learning rates, batch sizes, polynomial filter orders, number of Householder reflections, regularization weights, LR schedules, gradient clipping, and early stopping criteria are provided in the appendix hyperparameter tables.

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: All experiments report mean $\pm$ standard deviation over multiple seeds (3 for synthetic, 5 for traffic). Error bars capture variability across random initializations and, for synthetic experiments, across random manifold samplings. The synthetic experiment additionally validates empirical results against closed-form analytical projection distances, providing a strong correctness check.

   Guidelines:

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: In Appendix F, we indicate that the synthetic and traffic experiments have been run on a single (M4) CPU node and H200 GPU, respectively.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The research is foundational mathematical and theoretical work on geometric deep learning. It uses only publicly available benchmark datasets and poses no ethical concerns.

Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: The paper references a broader impact discussion in a dedicated appendix. The work is foundational in nature, extending Laplacian-based learning theory to infinite-dimensional signals, with potential positive applications in AI theory and methods, climate science, robotics, and sensor networks, and no foreseeable direct negative societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The paper introduces a theoretical framework and lightweight models for traffic forecasting. No pre-trained models, generative models, or scraped datasets with misuse risk are released.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: The METR-LA and PEMS-BAY datasets are properly cited (Li et al., 2018). These are standard public benchmarks widely used in the spatiotemporal forecasting community. Explicit license names are not stated, which is common practice for these datasets.

    Guidelines:

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

13. **New assets**

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

    Answer: [Yes]

    Justification: We release our code at `https://github.com/clabat9/hilbnet`, documented with a README and under the MIT license. No new datasets or pre-trained models are released.

    Guidelines:

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

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

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

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

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

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

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

    Answer: [N/A]

    Justification: The paper does not involve research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: The core methodology (Hilbert bundle theory, sheaf Laplacian convergence, HilbNet architecture) does not involve LLMs as any important, original, or non-standard component. LLMs have been used for writing, editing, and code assistance.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.