Dynamical Geometry and a Persistence *K*-Theory in Noisy Point Clouds

Sita Gakkhar(✉) and Matilde Marcolli

California Institute of Technology, Pasadena, CA 91125, USA
{sgakkhar,matilde}@caltech.edu

Abstract. The question of whether the underlying geometry of a dynamical point cloud is invariant is considered from the perspective of the algebra of trajectories of the points as opposed to their point-set topology. We sketch two approaches to identifying when the geometry remains invariant, one that accounts for a model of stochastic effects as well, and a second that is based on a persistence K-theory. Additional geometric structure in both approaches is made apparent by viewing them as finite noncommutative spaces (spectral triples) embedded inside the Hodge-de Rham spectral triple. A general reconstruction problem for such spaces is posed. The ideas are illustrated in the setting of understanding the dependence of grid cell population activity on environmental input.

Keywords: Grid cell modules · K-theory · Persistence homology · Noncommutative geometry · Stochastic differential geometry · Discrete differential geometry

1 Introduction

A dynamical point cloud is a family of point clouds (D_θ) parameterized by time or other environmental input, $\theta \in \Theta$. For each θ, the data, D_θ, are assumed to be sampled from a compact Riemannian manifold, M_θ. Characterizing the change in geometry and topology defined by the point cloud has important applications in many fields. Towards this, we study the geometry of a dynamic point cloud through discrete differential geometry and the persistence of the K_0 functor. This algebraic approach naturally connects with viewing the point clouds as finite spectral triples embedded inside the Hodge-de Rham spectral triple for M. The connection is provided by the results from [2] on the convergence of point cloud Laplacians to the Laplace-Beltrami operator and a Hodge theory on metric spaces developed by [1]. The point cloud Laplacians also allow for considering a stochastic version of the question with the Laplacian as the generator for the noise process. We begin by putting forward a model describing the case where the geometry is invariant over Θ up to stochastic effects and statistical testing in such a setup. Then we establish a stability theorem for an algebraic persistence theory to complement the topological persistence homology by capturing the dynamics of individual points without the complexity of multidimensional

F. Nielsen and F. Barbaresco (Eds.): GSI 2023, LNCS 14071, pp. 474–483, 2023.
https://doi.org/10.1007/978-3-031-38271-0_47

persistence. Finally, we consider the convergence of discrete Dirac operators for point clouds to the Dirac operator for the ambient Hodge-de Rham spectral triple. This is needed to be able to argue that the discretely sampled trajectories are sufficient to understand the geometry. A general reconstruction question is posed for such embedded finite spectral triples. The underlying motivation is understanding the modulation of grid cell firing by the environment. We start by introducing this illustrative example.

1.1 Modulation of Grid Cell Firing by Environmental Input

In the entorhinal cortex grid cells are cells with spatial firing fields organized on regular grids that form a part of the neural system responsible for navigation and mapping. Grid cells are organized in modules with structured correlations between different cells in the module. The neural code used by grid cell networks can be probed using persistence homology. In [10], Gardner et al, find that the activity of grid cell modules lies on a toroidal manifold that persists across brain states and offers support for continuous attractor models of grid cell activity. They also show evidence for environmental input-driven deformation of the geometry of population activity[1]. This can be thought of as an example of homeostatic plasticity. The question of the degree of stability of population dynamics is interesting, and one would like to relate this deformation to mechanistic models. A first step in this direction is putting forward a statistical test for the simplest case where the geometry is invariant and the point clouds evolve under a diffusion process on this fixed geometry.

We set this question up as follows: suppose that the spike train data from N neurons measured at K spatial locations $x_k^0, k \in [K], x_k \in \mathbb{R}^N$ in environmental conditions E_0 at time t_0. The environmental conditions are then updated to E_1 with firing data $(x_k^1)_{k\in[K]}$. The point cloud $\hat{M}_K^t := \{x_k^t : k \in [K]\} \subset \mathbb{R}^N$ changes with $t \in T$. The question now is of testing if the geometry M_K^t from which the point cloud $\hat{M}_K^t$ is invariant with respect to environmental input over $t \in T$, that is, $M_K^t = M_K^0 := M$, where the sample path of individual points, x_t^k, follows Brownian motion process on the invariant geometry M, that is, the diffusion generated by the Laplace-Beltrami operator, $\triangle_M$. While K is fixed, data from multiple runs of the experiment can be pooled to consider large size limit of the point cloud.

The choice of the process provides a natural null model for testing the presence of non-Markovian dynamics, as well as for testing synchronization in the point cloud. The hypothesis being tested is not just that the point cloud lives on an invariant geometry, that is, it's sampled from $M \times [0, T]$, but also that the time evolution follows Brownian motion on M. One can consider more general diffusion processes for such model testing, with the parameters learned from the time-series data, however, if the geometry is relevant then the Laplace-Beltrami operator is expected to play a role.

[1] [10, Tori persist despite grid distortions].

1.2 A Diffusive Model and Random Matrices

As a prelude to introducing L^2 Hodge theory [1], we consider the question of testing the hypothesis that the point cloud $M_K^t = M$ for all t. The Riemannian manifold (M, g), with $\dim M = d$ and metric g, is assumed to be embedded smoothly and isometrically in an ambient space, $\phi : M \hookrightarrow \mathbb{R}^N$, and for each $k \in K$, x_k^t is evolving by $\triangle_M/2$ diffusion on M.

Recalling that on a filtered probability space $(\Omega, F_*, \mathbb{P})$ a M-valued, F_*-adapted, stochastic process (X_t) is a (local-, semi-)-martingale on $[0, \tau)$ if $f(X_t)$ is a real-valued (local-, semi-)-martingale for all $f \in \mathcal{C}^\infty(M)$ where τ is a F_* stopping time (see, for instance, [12]). Brownian motion, $X := (X_t)$, on M is the $\triangle_M/2$ generated diffusion process, that is, a F_*-adapted process $X : \Omega \to W(M)$ (where $W(M)$ is the path space on M) such that for all $f \in \mathcal{C}^2(M)$, $\omega \in W(M)$, M^f as defined below is a local martingale:

$$M^f(\omega)_t := f(\omega_t) - f(\omega_0) - \frac{1}{2}\int_0^t \triangle_M f(X_s)ds \tag{1}$$

By the results of Belkin-Niyogi [2], the convergence of empirical estimates of Laplacians on finite metric space to $\triangle_M$ is known. This is formulated as follows: data $X_n = (x_i)_{i\in[n]}$ is n samples form M sampled with respect to uniform measure, μ_M, $\dim M = d$, giving an increasing sequence of metric spaces $X_1 \subset X_2 \subset \dots X_i \subset X_{i+1} \cdots \subset M$. To each X_n is the associated empirical Laplacian, $\triangle_{t_n,n}$, defined for $p \in M$ by $\triangle_{t_n,n} f(p) := \sum_{i\in[n]} K_{t_n}(p - x_i)(f(x_i) - f(p))nt_n^{d+2}$ where $K_{t_n}(u) = \exp(-\|u\|^2/4t_n)$ and t_n an appropriate sequence decreasing to 0, $\|\cdot\| = \|\cdot\|_{\mathbb{R}^N}$, c, then we have $\lim_{n\to\infty} \triangle_{t_n,n} f(x)/t_n(4\pi t_n)^{d/2} = \triangle_M f(x)/\mathrm{Vol}(M)$.

An analogous result holds for any probability measure μ_M on M. Now the local-martingale characterization of $\triangle_M/2$-diffusion (Eq. 1) applied to $f_i = \pi_i \circ \psi$, the coordinate functions of the smooth embedding ψ to easily test the question that $M_K^t = M$ for all t and X_k^t follows $\triangle_M/2$ diffusion. This is further simplified by noting that $f_i(X_s)$ is uniformly bounded and therefore a martingale, so the mean at each t is constant. The needed statistical test is just the test for constancy of the mean estimated by averaging data from l repeated experiments and using the control on $\triangle_M f(x)$ from [11] which gives a quantitative version of the convergence of the point cloud Laplacian. This is stronger than testing for stationary, e.g. using the unit root tests, as it's additionally required that the generator is the Laplacian.

Simplicial homology of random configurations and dynamical models for random simplicial complexes have been studied (for example, [6,8]), the simple example here suggests that (co)homology, both with rational coefficients and the α-scale theory of [1] for randomly evolving configurations is also meaningful from an applications perspective as well.

1.3 Discrete Differential Operators with Heat Kernel Weights

On a finite metric space, (X_n, d), with a probability measure μ, the point cloud Laplacian can be realized as Hodge Laplacian of a (co)chain complex [1]. Note that for a finitely supported measure ν on M, the point cloud Laplacian on M is an empirical estimate (via concentration bounds) for the functional approximation to the Laplace-Beltrami operator $\triangle_t f(x) = \int_X (f(x) - f(y))K_t(x,y)d\nu(y)$. We work in the picture that n point metric space X is n samples from M, d is the distance in ambient euclidean space, d_M the geodesic distance on M, and as n increases we have inclusions $i_n : X_n \to X_{n+1}$, $|X_n| = n$ and $X_{n+1} \setminus X_n$ is the one additional sample from M.

Fix $X_n = X$. Barthodi et al. [1] consider (co)chain complexes on $L^2(X^l)$ using the coboundary map $\delta_{l-1} : L^2(X^l) \to L^2(X^{l+1})$, $[\delta f](z_0, z_1 \dots z_l) = \sum_{i=0}^{l}(-1)^i \prod_{i\neq j} \sqrt{K(z_i,z_j)}\ f(z_0, \dots \hat{z}_i \dots z_l)$ where $X^l = \prod_{i\in[l]} X$, $L^\infty(X^2) \ni K : X^2 \to \mathbb{R}$ is symmetric, nonnegative and measurable; $K := K_t(\cdot,\cdot)$ is taken the t_n scaled heat kernel. The boundary map $\partial_l : L^2(X^{l+1}) \to L^2(X^l)$ is defined by $[\partial g](z_0 \dots z_{l-1}) = \sum_{i=0}^{l}(-1)^i \int_X \prod_{j=0}^{l-1} \sqrt{K(s,z_j)}$ $g(z_0 \dots z_{j-1}, s, z_{j+1} \dots z_{l-1})\ d\mu(s)$ and satisfies $\delta_{l-1}^* = \partial_l$, and the laplacian, $\triangle_l = (\delta_l^*\delta_l + \delta_{l-1}\delta_{l-1}^*)$ can be defined. The constructions and results also hold for $L_a^2(X^l) = \{f \in L^2(X^l) : f(x_0, \dots x_l) = (-1)^{\mathrm{sgn}(\sigma)} f(\sigma(x_0), \dots \sigma(x_l)), \sigma \in \mathcal{S}_{l+1}\}$. In [1], they also establish that for a Riemannian manifold, (X, g, μ), on restricting this construction to a suitable neighborhood of the diagonal, de Rham cohomology of X can be recovered and a Hodge decomposition exists for each $L^2(X^l)$.

Observing that $\triangle_0^t(f(x)) = \int_X (f(x) - f(y))K_t(x,y)d\mu(y)$, i.e., $\triangle_0|_{L^2(X)}$ is exactly the functional approximation to the Laplace-Beltrami operator which in the large sample-small t limit approaches the Laplace-Beltrami operator, and since on restricting to functions, Hodge-de Rham Laplacian agrees with the Laplace-Beltrami operator up to a sign suggests that in this limit $\delta^{(n)}$ associated to the sequence of n-point metric spaces (X_n) must approach the usual exterior derivative d acting on $\Omega^0(X)$. We give a quick proof using covariant Taylor series with respect to the canonical Riemannian connection ∇.

Theorem 1. *Suppose $U \subset \mathbb{R}^N$ is such that $M \cap U$ is a normal neighborhood of $x \in M$, and for any $y \in M \cap U$, $y \neq x$, $x(t)$ is the unique unit speed geodesic joining x, y, $v := \dot{x}(0)$. Then for $s = d_M(x,y)$ and $K_t(x,y) = \exp(-\|x-y\|_N^2/4t)$, $s = t + O(t^2)$ implies $|\delta f(x,y)/t - df_x(v)| = O(t)$.*

Proof. Since $x(t)$ is unit speed geodesic with $x(0) = x$, so $x(s) = y$. Expanding in a covariant Taylor series about $x(0)$, $f(x(t)) = \sum_{n=0}^{\infty} t^n/n! d^n/d\tau^n f(x(\tau))|_{\tau=0}$, with $d/d\tau = \dot{x}^i(\tau)\nabla_i$, gives $f(y) - f(x) = s \cdot df(v) + O(s^2)$ since first order term is $\dot{x}^i(\tau)\nabla_i f|_{\tau=0} = s \cdot g(v, \nabla f(x)) = s \cdot df_x(v)$. We have $\delta f(x,y) = \sqrt{K_t(x,y)}(f(y) - f(x)) = \sqrt{K_t(x,y)} s\dot{d}f_x(v) + \sqrt{K_t(x,y)}O(s^2)$. For fixed x, using that there exists $\eta \geq 0$, such that $d_M(x,y)^2 - \|x-y\|_N^2 = \eta(y)$ with $|\eta(y)| \leq C d_M(x,y)^4$ for a constant C on the normal neighborhood U, so $\|x-y\|_N^2 = d_M(x,y) - \eta(y)$. Using $e^\alpha = 1 + O(\alpha e^\alpha)$ for $\alpha > 0$, $1/(1+\alpha) \leq 1 + O(\alpha)$ yields the following estimate from which the result follows for $s = t + O(t^2)$

$$
\begin{aligned}
\left|\sqrt{K_t(x,y)}\frac{s}{t}df(v) - df(v)\right| &= \left|\left(e^{\eta(y)}e^{-d_M(x,y)^2/8t}\frac{s}{t} - 1\right)df(v)\right| \\
&\leq \left|\left(\frac{s}{t}(1 + O(s^2/t))(1 + O(s^4/t)) - 1\right)df(v)\right|
\end{aligned}
$$

In the large sample limit as the sampled points get closer s/t approaches identity while $s^k/t, k > 1$ terms vanish, and the exterior derivative is recovered. This observation is the basis for the attempt in Sect. 3 to formalize how sample paths, x_k^t, (from Sect. 1.1) encode the underlying geometry using Hodge-de Rham spectral triples. To warm up to the idea of replacing topological spaces (X_n) by the algebras $\mathcal{C}(X_n)$, we consider the persistence theory K_0 functor and use it towards analyzing dynamical geometry in point clouds.

2 $\mathbb{Q} \otimes K_0$-Persistence

Dynamical point clouds have been studied through persistent homology theories that use multiple persistence parameters for the incomparable space and time dimensions [13]. However, theories that use independent persistence parameters introduce complexity that intuitively is not necessary. Consider the question of detecting synchronization. Suppose in the extreme case, the point cloud completely synchronizes to evolve by rotation, so that the distance matrices $[D_{ij}]_{i,j\in[K]}$, are invariant in time, and persistence homology is constant for every value of space and time persistence parameters. One can detect this synchronization by analyzing the time persistence, but one now needs to test ranges of multiple independently varying persistence parameters to assign statistical confidence.

Since in the setup of the basic question, we are not exploring the development of new structures in relationships between points in time and are only interested in the sample paths of the points themselves, one expects that persistence in time is unnecessary. This intuition can be verified by showing that a persistence theory with only spatial parameters is sufficient in this setting. Furthermore, this theory is shown to be equivalent to a topological persistence theory.

2.1 A Category-Theoretic Formulation of Persistence

In [3] Bubenik and Scott formulate persistence homology abstractly in terms of functor F from a small poset category C into a category D called C-indexed diagram in D. The space of such functors with natural transformations is the category D^C. Composing a diagram in the category of topological spaces $\textsc{Top}$ indexed by $(\mathbb{R}, \geq)$, $F \in \textsc{Top}^{(\mathbb{R},\geq)}$, $F : (\mathbb{R}, \geq) \to \textsc{Top}$ with the k-th homology functor H_k into the category of finite dimensional vector spaces $\textsc{Vec}$ gives a diagram $H_k F \in \textsc{Vec}^{(\mathbb{R},\geq)}$. For a topological space X, a map $f : X \to \mathbb{R}$ defines a functor $F \in \textsc{Top}^{(\mathbb{R},\geq)}$ by $F(a) = f^{-1}((-\infty, a])$, and from this data the p-persistent k-th homology group for the topological space X is defined as the image of map $H_k F(a \leq a+p)$ induced on homology by the inclusion $H_k F(a) \hookrightarrow H_k F(a+p)$. The construction of a persistence K-theory is analogous. We first use the functor $\mathcal{C} : \textsc{Top} \to \mathcal{C}_1^*$, where $\mathcal{C}_1^*$ is the category of unital $\mathcal{C}^*$ algebras, that associates to compact Hausdorff topological spaces X, Y, the unital $\mathcal{C}^*$-algebras $\mathcal{C}(X), \mathcal{C}(Y)$ and to continuous map $\phi : X \to Y$, the pullback, $\phi^* : \mathcal{C}(Y) \to \mathcal{C}(X), \phi^*(h) = h \circ \phi$. Note that $\mathcal{C}$ reverse the direction of the arrows: for $\epsilon > 0$, the inclusion $i : F(a) \hookrightarrow F(a+\epsilon)$ induces $i^* : \mathcal{C}(F(a+\epsilon)) \to \mathcal{C}(F(A))$, we adjust this by using the opposite category to index, equivalently the diagram $F : (\mathbb{R}, \geq) \to \textsc{Top}$ the associated diagram is $\mathcal{C}F : (\mathbb{R}, \geq) \to \mathcal{C}_1^*$, $-a \to \mathcal{C}(F(a))$.

2.2 The $\mathbb{Q} \otimes K_0$-Functor: Stability and Computation

On diagrams $F_1, F_2 \in D^C$, there exists an extended pseudo-metric, d^{IL}, defined as $d^{IL}(F_1, F_2) = \min\{\epsilon : \epsilon > 0, F_1, F_2 \text{ are } \epsilon \text{ interleaved}\}$ where F_1, F_2 are ϵ-interleaved if there exists natural transformations $\phi_{12} : F_1 \Rightarrow F_2, \phi_{21} : F_2 \Rightarrow F_1$ such that the following diagrams commute for $i, j \in \{1, 2\}, i \neq j$, the horizontal arrows being the inclusions of the diagram:

$$\begin{tikzcd} F_i(a) \arrow[r] \arrow[d, "\phi_{ij(a)}"] & F_i(b) \arrow[d, "\phi_{ij}(b)"] \\ F_j(a+\epsilon) \arrow[r] & F_j(b+\epsilon) \end{tikzcd} \qquad \begin{tikzcd} F_i(a) \arrow[r] \arrow[dr, "\phi_{ij}(a)"'] & F_i(a+2\epsilon) \\ & F_j(a+\epsilon) \arrow[u, "\phi_{ji}(a+\epsilon)"'] \end{tikzcd}$$

The K_0-functor is the functor from $\mathcal{C}_1^*$ to the category of abelian groups $\textsc{AbGrp}$ that associates to an unital $\mathcal{C}^*$-algebra its Grothendieck group. We consider the diagrams in $\textsc{AbGrp}^{(\mathbb{R},\geq)}$, $K_0\mathcal{C}F$. The p K_0-persistence is now defined for the diagram $\mathcal{C}F$ as the image of map $K_0F_{\mathcal{C}}(a \geq a+p)$ induced on K_0-group by the map $K_0\mathcal{C}F(a) \to K_0\mathcal{C}F(a+p)$. As for topological persistence, a stability theorem is needed that ensures that similar topological spaces have similar K_0 persistence for their continuous function algebras. We have that $\mathcal{C}$ is contractive with respect to the interleaving distance even though it reverses the arrows. And since by [3, Prop 3.6], for any functor $H : \mathcal{C}_1^* \to E$ to any category $d^{IL}(H\mathcal{C}F_1, H\mathcal{C}F_2) \leq d^{IL}(\mathcal{C}F_1, \mathcal{C}F_2)$. This yields the needed stability theorem analogous to [3, Thm 5.1] as a corollary.

Lemma 1. *For $F_1, F_2 \in \textsc{Top}^{(\mathbb{R},\geq)}$, $d^{IL}(\mathcal{C}F_1, \mathcal{C}F_2) \leq d^{IL}(F_1, F_2)$*

Proof. This follows since if F_1, F_2 are ϵ-interleaved then $\mathcal{C}F_1, \mathcal{C}F_2$ are as well: the associated natural transformation obtained by composing $\phi_{ij} \circ \mathcal{C}$ and the as $\mathcal{C}$ simply reverse the arrows the interleaving relations still hold.

Corollary 1. *If $F_1, F_2 \in \textsc{Top}^{(\mathbb{R},\geq)}$ are such that $F_i(a) = f_i^{-1}((-\infty, a])$, then*

$$d^{IL}(K_0\mathcal{C}F_1, K_0\mathcal{C}F_2) \leq d^{IL}(\mathcal{C}F_1, \mathcal{C}F_2) \leq \|f_1 - f_2\|_\infty$$

Proof. From the proof of [3, Thm 5.1], $d^{IL}(F_1, F_2) \leq \|f_1 - f_2\|_\infty$, and the rest follows.

For increasing finite metric spaces arising by sampling from a manifold M, $X_1 \hookrightarrow X_2 \ldots \hookrightarrow M$, the inclusions $X_n \hookrightarrow X_{n+1}$ induce maps $\mathcal{C}(X_{n+1}) \to \mathcal{C}(X_n)$. Recovering the algebra $\mathcal{C}(M)$ in large n limit of such systems is difficult as projective limits of $\mathcal{C}^*$-algebras are more general pro $\mathcal{C}^*$-algebras. Even K_0 may not be continuous under the projective limits. Keeping in mind that the goal is simply a statistical test for the invariance of the underlying geometry, one can use the following observation to derive the test.

Lemma 2. $\mathbb{Q} \otimes K_0(\mathcal{C}(X)) \otimes \cong H^{even}(X, \mathbb{Q})$

Proof. This is obvious from results in topological K-theory [16] : $K^0(X) \otimes \mathbb{Q} \cong H^{\text{even}}(X, \mathbb{Q})$ for any topological space X where $K^0(X)$ is topological K^0 group associated to isomorphism classes of vector bundles over X. When X is compact Hausdorff space, as abelian groups $K_0(\mathcal{C}(X)) \cong K^0(X)$, and on taking the tensor product with $\mathbb{Q}$, they are isomorphic as $\mathbb{Q}$-vector spaces.

This reduces the algebraic K-theoretic persistence to the persistence of the even rational cohomology of the topological space X for which the sample paths approximate $\mathcal{C}(X)$. We offer a candidate space next such that the a topological persistence parameter can be obtained from K-persistence parameter.

Notice that if the time evolution is constrained to be by a possibly random isometry, then the hypothesis that the geometry of the point cloud is invariant translates to the null model being that the time evolution of the topological Rips simplicial complex at persistence parameter ϵ is simply the mapping cylinder M_1, formed by gluing $(x_k^t, t) \sim (x_k^{t+1}, t+1)$. Since the evolution is isometric, the maps are simplicial under the null hypothesis, and l-cells in complexes, $X_\epsilon^t, X_\epsilon^{t+1}$, at times $t, t+1$, can be glued. Confidence in how well the true data conforms to the hypothesis can be quantified by testing the cohomology of the time-evolved complex X^T for actual data against the expected.

If the evolution is not isometric, then picking a single persistence parameter is difficult as distances in various parts of the geometry will change differently. This can be accounted for by using that as in the Brownian motion diffusive model, the generator is the Laplace-Beltrami operator, $\triangle$, which is being approximated by the point cloud Laplacian, $\triangle^{\text{PC}}$, the evolution will be isometric in expectation after adjusting for the eigenvalues of $\triangle^{\text{PC}}$; we will work with this rescaled metric. The rescaling does not affect the cohomology and allows for using a uniform spatial persistence parameter for the time-evolved complex. The actual data can now be tested against the simulated data or against the expectation to see if the null hypothesis of a diffusive model can be accepted.

The presence of stochastic effects is measured by the distribution of lifetimes of the simplices in this process since if the data is not evolving by a process generated by the Laplacian, then rescaling by the eigenvalues of the point cloud Laplacian will not yield isometric evolution, leading to simplices splitting and merging. At the same time, longer than expected lifetimes for simplices for the unscaled metric indicate likely synchronized sub-populations and possible homeostatic plasticity in the population response to input, which is of interest.

3 Embedded Finite and Hodge-de Rham Spectral Triples

For T large, the data of Brownian motion sample paths $\gamma : [0, T] \to M$ on a finite point cloud, composing with coordinate functions of the embedding $\psi : M \hookrightarrow \mathbb{R}^N$ gives a discretized version of the algebra $\mathcal{C}(M)$ because of the asymptotics of the time taken to get within r of each point, the r-covering time [7]. If this is enough to recover the geometry of M is central to the program we have outlined. This is best viewed as a question in noncommutative geometry: we recall how commutative geometry is encoded in the noncommutative language. The Hodge-de Rham spectral triple, $\mathfrak{A}_M$, for Riemannian manifold (M, g) is the data $(\mathcal{C}^\infty(M), \Omega^\bullet(M), d+d^\dagger)$ where $d+d^\dagger$ is the Hodge-de Rham Dirac operator, d the exterior derivative on differential forms $\Omega^\bullet(X)$, $d^\dagger$ the Hodge dual. By Connes' spectral characterization of manifolds [4], (M, g) can be recovered from $\mathfrak{A}_M$. A finite spectral triple is the triple, $\mathfrak{A}_F := (\mathcal{A}_F, H_F, D_F)$, where $\mathcal{A}_F$ is an unital $*$-algebra represented faithfully on a Hilbert space $H_F, \dim H_F$ finite, and D a symmetric operator on H_F subject to some additional requirements. There's a standard representation of a finite metric space as a finite spectral triple.

We instead define an alternative representation using theorem 1 to obtain a finite Dirac operator. Suppose $\dots X_i \subset X_{i+1} \dots$ is an increasing sequence of metric space sampled from M, with $X_n = \{x_i : i \in [n]\}$. Then using the L^2-Hodge theory, with a uniform measure on X_n (except weighed multiply if $x_i = x_j, i \neq j$), we associate to it the restriction of the algebra $\mathcal{C}^\infty(M)$ and $\Omega^\bullet(M)$. $\Omega^\bullet(M)$ is permissible as the space of co-chains is alternating, that is, $L_a(X_n^\bullet)$. Similar to theorem 1 it's possible to show that for the operator $\delta_{l-1}^{(n)}$ on $L^2(X^l)$, $\delta f(x_0 \dots x_l)$ converges to $df_{x_0}(v_1 \dots v_l)$ where v_i is the tangent at x_0 to the unit speed geodesic to x_i, the idea being to fix $l-1$ of x_i's to get back to 1-cochain setting, although it needs to be checked that this is well defined regardless of order and number of fixed x_i's. This can be achieved using continuity of f as in the limit we restrict to infinitesimal neighborhoods of x_0. From this, the result below follows which for transparency can be roughly stated as –

Theorem 2. *The finite Dirac operators, $D_n := \delta^{(n)} + (\delta^{(n)})^*$, for X_n converges to the Hodge-de Rham Dirac operator $d + d^\dagger$ for M.*

To reconstruct the full Hodge-de Rham spectral triple from finite spectral triples (and (M, g) by [4]) the knowledge of $\mathcal{C}^\infty(M)$ and $\Omega^\bullet(M)$ cannot be assumed. For recovering the algebra of the spectral triple, instead of taking the projective limit of $\mathcal{C}(X_n)$, we use a classical result in PL-topology [14]: M being smooth implies there exists a homeomorphism $\phi : K \to M$, where K is a polyhedron with triangulation $\{\sigma_i\}$ and ϕ is a piecewise diffeomorphism on σ_i, and therefore, $\mathcal{C}(M) \cong \mathcal{C}(K)$.

For the polyhedron K, viewed as the geometric realization $|\Sigma|$ of an abstract simplicial complex Σ on the finite vertex set $V_\Sigma = i \in [N]$ for $\{\sigma_i\}$, define $\mathcal{C}_\Sigma^{ab}$ as the abelianization of the universal $\mathcal{C}^*$-algebra generated by positive generators $h_i, i \in V_\Sigma$, $h_{i_1} h_{i_2} \dots h_{i_k} = 0$ whenever $\{i_j : j \in [k]\} \subset \Sigma$ and for all $m \in V_\Sigma, \sum_{k \in V_\Sigma} i_m i_k = i_m$ with the dense subalgebra generated algebraically on the same generators and relations. Then from [5], $\mathcal{C}_\Sigma^{ab} \cong \mathcal{C}_0(|\Sigma|)$ where $|\Sigma|$ is the geometric realization of Σ. As M, K are compact, $\mathcal{C}_\Sigma^{ab} \cong \mathcal{C}(M)$. The last ingredient needed to recover the Hodge-de Rham spectral triple is how the Dirac operator acts on $\mathcal{C}(K)$, but this is given by the homeomorphism ϕ, although some care is required as ϕ is only a piece-wise diffeomorphism (so the action of Dirac operator is not everywhere defined and we need to restrict to a differentiable subalgebra).

Finally, from d and $\mathcal{C}^\infty(M)$, $\Omega^\bullet(M)$ can be constructed. The Dirac operators for the finite spectral triples we have used are weighed by the Euclidean heat kernel of the ambient space and are not standard finite spectral triples. This spectral triple with the Dirac operator coming from the L^2 Hodge theory is defined as an *embedded* finite spectral triple. The details of convergence to the Hodge-de Rham spectral triple[2] are developed in forthcoming work [9].

We end this article by posing the question of computationally reconstructing the Hodge-de Rham spectral triple, that is, recovering K and ϕ from the point cloud data, (X_n), in the large n limit. In particular, one does not expect to have access to the Euclidean embedding $\psi : M \to \mathbb{R}^N$, but can only construct the simplicial complex from sampled points, and the discrepancy of the action of Dirac operator on constructed simplex and M needs to be bound in terms of the geometry (e.g. M's maximum sectional curvature).

References

1. Bartholdi, L., Schick, T., Smale, N., Smale, S.: Hodge theory on metric spaces. Found. Comput. Math. **12**, 1–48 (2012)
2. Belkin, M., Niyogi, P.: Towards a theoretical foundation for Laplacian-based manifold methods. J. Comput. Syst. Sci. **74**(8), 1289–1308 (2008)
3. Bubenik, P., Scott, J.A.: Categorification of persistent homology. Discrete Comput. Geom. **51**(3), 600–627 (2014)
4. Connes, A.: On the spectral characterization of manifolds. J. Noncommutative Geom. **7**(1), 1–82 (2013)
5. Cuntz, J.: Noncommutative simplicial complexes and the Baum-Connes conjecture. Geom. Funct. Anal. GAFA **12**(2), 307–329 (2002)
6. Decreusefond, L., Ferraz, E., Randriambololona, H., Vergne, A.: Simplicial homology of random configurations. Adv. Appl. Probab. **46**(2), 325–347 (2014)
7. Dembo, A., Peres, Y., Rosen, J.: Brownian motion on compact manifolds: cover time and late points. Electron. J. Probab. **8** (2003)
8. Fountoulakis, N., Iyer, T., Mailler, C., Sulzbach, H.: Dynamical models for random simplicial complexes. Ann. Appl. Probab. **32**(4), 2860–2913 (2022)
9. Gakkhar, S., Marcolli, M.: Embedded finite and Hodge-de Rham spectral triples (2023, in preparation)
10. Gardner, R.J., et al.: Toroidal topology of population activity in grid cells. Nature, 1–6 (2022)
11. Giné, E., Koltchinskii, V.: Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. IN: Lecture Notes-Monograph Series, pp. 238–259 (2006)
12. Hsu, E.P.: Stochastic Analysis on Manifolds, No. 38, American Mathematical Soc. (2002)

[2] The very recent article [15], considers a similar approach for convergence of finite spectral triples without the heat kernel weights, building on the Hodge theory for the non-commutative laplacian. Using the L^2 Hodge theory complements their ideas, and in particular, gives a canonical choice for both the Dirac operator and Laplacian regardless of the distribution on the finite spaces, making progress on an open research direction they cite.

13. Kim, W., Mémoli, F.: Spatiotemporal persistent homology for dynamic metric spaces. Discrete Comput. Geom. **66**, 831–875 (2021)
14. Lurie, J.: Topics in geometric topology (18.937) (2009)
15. Tageddine, D., Nave, J.C.: Statistical fluctuation of infinitesimal spaces. arXiv preprint arXiv:2304.10617 (2023)
16. Wegge-Olsen, N.E.: K-theory and C^*-algebras (1993)