Research Statement
Building theoretical foundations of applied topology

Tony Xiaochen Xiao

My research is in applied/computational topology and mathematical foundations of data science. I am particularly interested in the interplay between *algebraic/geometric topology, algorithms, and applications* in science and engineering.

1. Introduction and Overview

A fundamental challenge in topological data analysis is transforming discrete topological invariants—such as persistence barcodes—into continuous, interpretable geometric features for data. While persistent homology has matured as a diagnostic tool for detecting topological structure in data, the question of **how to explicitly represent these features as coordinate functions** remains largely open beyond dimension one.

My research seeks to address this challenge through the systematic construction of **Eilenberg-MacLane coordinates**—maps $f : X \to K(G, n)$ from data to Eilenberg-MacLane spaces—for arbitrary G (Abelian) and n that provide canonical geometric realizations of persistent cohomology classes. In my dissertation, supervised by Professor Jose A. Perea, I have developed a theoretical framework and computational algorithms for constructing these coordinates across all dimensions $n \geq 1$ and arbitrary discrete Abelian coefficient groups G. This is a data science implementation of Brown representability: For a CW complex B and an Abelian group G, there is an one-to-one correspondence $H^n(B; G) \cong [B, K(G, n)]$. This work generalizes previous results on circular coordinates (De Silva, Morozov, Vejdemo-Johansson[1][2]; Perea[3]), projective coordinates (Perea[4]), lens coordinates (Polanco, Perea[5]), and toroidal coordinates (Scoccola et al.[6])—all limited to specific low-dimensional cases—to a unified framework with broader applicability. My contributions span theory (constructive proofs via soft sheaves), algorithms (polynomial-time recursive procedures), and computation (implemented and validated on synthetic datasets), positioning this work at the intersection of pure algebraic topology and practical data science.

2. Dissertation Research: Completed Work

2.1. Main Theoretical Results

Theorem 2.1. *Let X be a finite set embedded in a metric space $(\mathbb{X}, d)$ with Vietoris-Rips filtration $\mathcal{R}(X)$, and let G be a discrete Abelian group. For each persistent cohomology class $[\eta] \in PH^n(\mathcal{R}(X); G)$ with birth/death times b, d (suitably defined) satisfying $d > 2b$, there exists a canonical continuous map $f_\eta : X^{(\alpha)} \to K(G, n)$ defined on the α-offset $X^{(\alpha)}$ for any $\alpha \in (b, d/2)$. Moreover, the assignment $[\eta] \mapsto [f_\eta]$ defines an injection*

$$PH^n(\mathcal{R}(X); G) \to [X^{(\alpha)}, K(G, n)]$$

Key Innovation: My proof is constructive, providing explicit formulas via the inverse of the Čech coboundary operator's connecting homomorphism. This yields a canonical choice of coordinate map—a concrete geometric object that can be computed and analyzed—rather than merely proving existence up to homotopy equivalence. Previous work established similar correspondences but only for specific (G, n) pairs and without the systematic framework presented here. The following table illustrates how my framework unifies and extends existing coordinate constructions:

Coordinate Type	Cohomology Source	Target Space	Author
Circular	$PH^1(\mathcal{R}(X); \mathbb{Z})$	$K(\mathbb{Z}, 1) = S^1$	De Silva et al. [1][2]; Perea [3]
Torodial	$PH^1(\mathcal{R}(X); \mathbb{Z}^l)$	$K(\mathbb{Z}^l, 1) = T^l$	Scoccola et al. [6]
Real Projective	$PH^1(\mathcal{R}(X); \mathbb{Z}_2)$	$K(\mathbb{Z}_2, 1) = \mathbb{RP}^\infty$	Perea [4]
Complex Projective	$PH^2(\mathcal{R}(X); \mathbb{Z})$	$K(\mathbb{Z}, 2) = \mathbb{CP}^\infty$	Perea [4]
Lens	$PH^1(\mathcal{R}(X); \mathbb{Z}_q)$	$K(\mathbb{Z}_q, 1) = L_q^\infty$	Polanco, Perea [5]
My Framework	$PH^n(\mathcal{R}(X); \mathbb{Z}_q)$	$K(G, n) = B^n G$	All $n \geq 1$, any f.g. Abelian G

Technical Framework: I establish this correspondence through a chain of isomorphisms connecting Čech cohomology—taken with coefficients in the sheaf $\mathcal{F}_G$ of continuous G-valued maps—and induced by the universal bundle $G \to EG \to BG$:

$$\check{H}^n(B, \mathcal{F}_G) \cong \check{H}^{n-1}(B, \mathcal{F}_{BG}) \cong \ldots \cong \check{H}^1(B, \mathcal{F}_{B^{n-1}G}) \cong \mathrm{Prin}_{B^{n-1}G}(B) \cong [B, B^n G] = [B, K(G, n)]$$

Each isomorphism is proven constructively using (i) soft sheaf cohomology vanishing theorems ($\check{H}^n(B; \mathcal{F}_E) = 0$ for contractible E), (ii) principal bundle classification relating Čech cohomology to fiber bundles, and (iii) recursive gluing procedures via partitions of unity.

2.2. Algorithmic Contributions

My constructive proof directly yields a systematic computational procedure: the Sparse Eilenberg-MacLane Algorithm. Given a persistent cohomology class $\eta \in PH^n(\mathcal{R}(X); \mathbb{Z}/q)$ with a prime q, the algorithm determines a natural choice of coefficient group G and, with respect to this G, computes the corresponding coordinate function $f_\eta : X^{(\alpha)} \to K(G, n)$. The following steps give the framework:

1. Landmark sampling: Select N landmarks $L \subseteq X$ reducing computation from $\#(X)$ data points to $N << \#(X)$ landmarks.

2. Coefficient lifting: Lift the detected cohomology class from $\mathbb{Z}/q$ coefficients to an appropriate discrete Abelian group G using the Bockstein connecting homomorphism. This determines the most informative coefficient system—for instance, distinguishing whether a class lifts to $\mathbb{Z}$-coefficients or remains torsion ($\mathbb{Z}/(q^k)$-coefficients)—thereby identifying the precise target space $K(G, n)$.

3. Recursive cohomology-degree reduction: Iteratively compute the inverse operators $\check{Q}^n(\eta) \mapsto \check{Q}^{n-1}(\check{Q}^n(\eta)) \mapsto \ldots \mapsto \check{Q}^1(\ldots(\check{Q}^n(\eta)))$, where each $\check{Q}^k$ inverts the connecting homomorphism Δ while passing to the iterated classifying space $B^{n-k+1}G$ at the next level.

$$
\begin{tikzcd}[column sep=small]
\ldots \arrow[r] & \check{H}^{k-1}(B, \mathcal{F}_{B^{n-k}G}) \arrow[r] & \check{H}^{k-1}(B, \mathcal{F}_{EB^{n-k}G}) \arrow[r] & \check{H}^{k-1}(B, \mathcal{F}_{BB^{n-k}G}) \arrow[r, "\Delta"] & \check{H}^{k}(B, \mathcal{F}_{B^{n-k}G}) \arrow[l, bend left, dashed, "\check{Q}^k"] \arrow[r] & \ldots
\end{tikzcd}
$$

In terms of coding, $\check{Q}^k$ can be assembled into a row-by-row calculation on an $\underbrace{N \times N \times \cdots \times N}_{k \text{ times}}$ $EB^{n-k}G$-valued matrix/tensor, where N denotes the cardinality of the open cover of B. For example, the following is a procedure for calculating $\check{Q}^1$ on an open cover $\{U_0, U_1, U_2\}$, with the context explained in experiment 1.

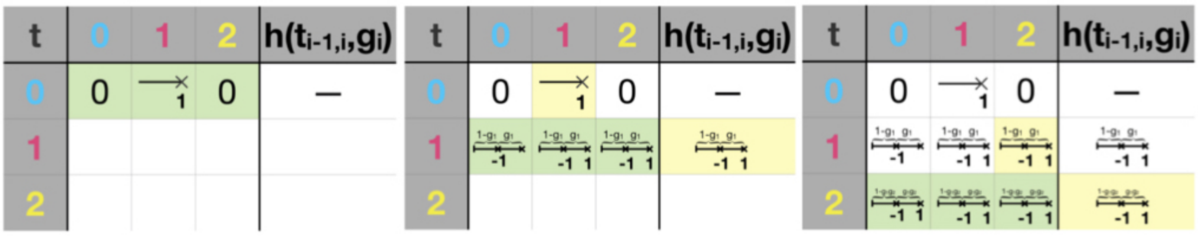

Figure 1: An illustration of $\check{Q}^1$—how the Eilenberg-MacLane Algorithm works at $n = 1$

4. Coordinate assembly: Build $f_\eta : X^{(\alpha)} \to B^nG$ by gluing local transition functions via a partition of unity subordinate to the landmark cover.

2.3. Experimental Validation

I have implemented the pipeline and validated the framework through three experiments on synthetic datasets:

Experiment 1 (Validation): Computing $K(\mathbb{Z}, 1)$-coordinates for data sampled from S^1 reproduces results consistent with De Silva's [1] and Perea's circular coordinates [3] up to homotopy, confirming algorithmic correctness. With a harmonic representative chosen, it would output a smooth circular coordinate.

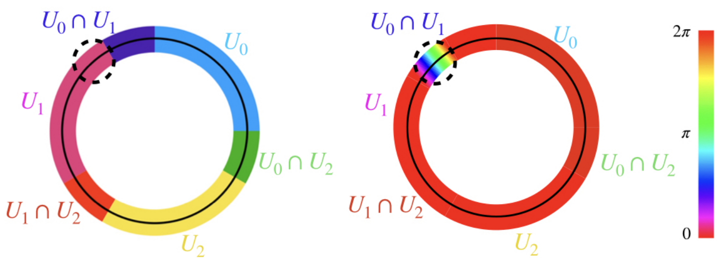

Figure 2: The space $X = S^1$ covered by three open sets $\{U_0, U_1, U_2\}$, and its Eilenberg-MacLane coordinates $f_\eta : X \to S^1$ derived from $[\eta]$, the generator of $H^1(\mathcal{N}(\mathcal{U}); \mathbb{Z})$, is labeled by the color map.

Experiment 2 (Dimensional Scalability): Computing $K(\mathbb{Z}, 2)$-coordinates for data sampled from S^2 demonstrates successful iteration of the inverse Čech coboundary operator across dimensions $2 \to 1 \to 0$, confirming applicability to arbitrary n. Results exhibit a structure analogous to complex projective coordinates, visualized via projective principal component analysis.

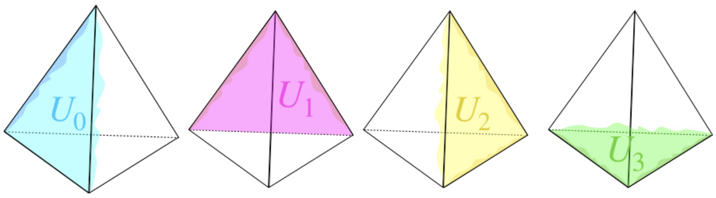

Figure 3: The space $X = S^2$ covered by four open sets $\{U_0, U_1, U_2, U_3\}$. The Eilenberg-MacLane coordinate is: $\forall x \in X, f_\eta(x) := \begin{cases} 0 & \text{if } x \in U_0 \\ (-1, (1-g_1(x), g_1(x))) & \text{if } x \in U_1 - U_0 \\ (-1, (1-g_1(x)g_2(x), g_1(x)g_2(x))) & \text{if } x \in U_2 - U_0 \cup U_1 \end{cases}$, where $g_1, g_2 : X \to [0,1]$ are some defined bump functions.

Experiment 3 (Novel Examples): I present initial computational examples of $K(\mathbb{Z}_2, 2)$-coordinates—absent from prior literature—which illustrate the framework's capability of capturing genuinely new topological features beyond one-dimensional persistence cohomology. These examples represent a promising direction for ongoing development.

2.4. Presentations and Dissemination

Conference Presentations: 30-minute talk: Young Topologists Meeting 2025 (Stockholm, June 2025); 1-hour talk: Brandeis Graduate Student Seminar (Brandeis, Oct 2023).

Lightning talks & posters: Applied Algebraic Topology Research Network (AATRN, Chicago, August 2025); Topology of Arrangements with an Eye to Applications (Pisa, September 2025); the 4th annual Centre for Topological Data Analysis conference (Oxford, August 2024); Applied Topology Montreal Summer School (Montreal, June 2024); Mid-Atlantic Topology Conference 2024 (Boston, March 2024).

Publications: In preparation: "Extracting Sparse Eilenberg-MacLane Coordinates via Soft Sheaves".

3. Future Research Directions

My future research develops along three interconnected threads, balancing theoretical depth with practical applications.

3.1. Near-Term: Stability Theory (Years 1-2)

Motivation: Persistent homology satisfies stability theorems with respect to perturbations in input data (Chazal et al.[7]). However, the stability of the coordinate functions themselves remains an open problem. Establishing such results is essential for applying these methods to noisy real-world data. I will prove that Eilenberg-MacLane coordinates vary continuously under data perturbations. Specifically, I aim to establish:

Conjecture (Stability of Coordinates): If persistent cohomology classes $[\eta], [\eta']$ from datasets X, X' satisfy bottleneck distance $d_B([\eta], [\eta']) < \varepsilon$, then the corresponding coordinate functions f_η, f'_η satisfy a Lipschitz-type bound on appropriate function spaces.

3.2. Near-Term: Software and Computational Optimization (Years 1-3)

I will transform my algorithms into accessible, optimized software tools for the TDA community. My development plan includes exploiting the sparsity structure of Čech cochain matrices for large-scale computations ($N > 1000$ landmarks), developing faster approximate coordinate computation with theoretical error bounds, and creating a Python package EMLCoord with a scikit-learn compatible API that interfaces with existing TDA libraries (DreiMac [8]). This will enable practitioners without deep topology expertise to apply these methods. I am eager to collaborate with research software engineers and computational topologists to optimize implementation and ensure community adoption.

3.3. Medium-Term: Applications and New Theoretical Directions (Years 2-4)

Direction 1: Neuroscience Applications: Brain functional connectivity networks from fMRI data naturally exhibit higher-order structure beyond pairwise correlations. I will apply $K(G, 2)$-coordinates to detect and quantify higher-order synchronization patterns in neural activity, topological biomarkers for neurological conditions, and multi-scale organization of brain networks. This is a natural extension of existing TDA applications to neuroscience (Sizemore et al.[9], Petri et al.[10]), and I am open to establishing collaborations with neuroscience groups during my postdoc to access datasets and domain expertise.

Direction 2: Multiparameter Persistent Cohomology: An open question is whether Eilenberg-MacLane coordinates can extend to multiparameter persistent cohomology, where modules are defined over polynomial rings $k[t_1, ..., t_d]$. I will investigate whether soft sheaf techniques apply to multigraded structures, starting with two-parameter cases ($d = 2$) as proof-of-concept and connecting to recent work on multiparameter persistence theory (Kalisnik[11], Oudot[12]). This is a high-risk, high-reward direction: success would provide the first coordinate-based framework for multiparameter TDA, opening new research directions, while even partial results for special cases would be valuable.

Direction 3: Persistent Geometric Structures: Beyond individual coordinate functions, I am interested in developing a theory of persistent geometric structures that addresses how coordinates in different dimensions ($K(G, n), K(G, m)$) interact, whether we can define "persistent principal bundles" capturing multi-scale geometric features, and what relationships exist to persistent homology with local coefficient systems. This connects my work to recent developments in categorical and sheaf-theoretic approaches to TDA (Bubenik et al.[13][14], Curry[15]).

4. Broader Impact and Research Philosophy

Impact on TDA Practice: My research transforms discrete topological invariants (barcodes) into continuous, interpretable coordinate representations. This enables:

(1) Topological dimensionality reduction: Eilenberg-MacLane coordinates provide a principled way to embed high-dimensional data into geometrically meaningful spaces while preserving global topological structure, which can be used for nonlinear dimensionality reduction.

(2) Machine learning integration: Continuous features compatible with standard ML pipelines (regression, classification, clustering).

(3) Visualization: Projection of high-dimensional coordinates for human interpretation.

Impact on Pure Mathematics: My constructive treatment of the classical representability theorem $H^n(B;G) \cong [B, K(G,n)]$ contributes to computational algebraic topology, demonstrating how abstract existence results can be made algorithmic.

Research Philosophy: I am guided by three principles:

1. Constructive mathematics: Every existence theorem should yield an explicit algorithm. My dissertation exemplifies this—each isomorphism in the theoretical chain translates directly to a computational procedure.

2. Systematic generalization: Rather than isolated results for special cases, I develop unified frameworks. My treatment of all dimensions $n \geq 1$ and arbitrary coefficient groups G demonstrates this approach.

3. Bridging theory and practice: I actively engage with both pure mathematics (for theoretical rigor) and applied problems (for relevance). My planned applications to neuroscience and ongoing software development reflect this commitment.

5. Vision for a Postdoctoral Research Program

My postdoctoral research will solidify the theoretical foundations of Eilenberg-MacLane coordinates while expanding into new applications and theoretical directions. The near-term goals (stability theory, software optimization) will establish my work as a reliable tool for TDA practitioners. The medium-term directions (applications, multiparameter extensions) will demonstrate versatility and open new research questions. Together, these threads position me to build an independent research program that contributes both theoretical depth and practical impact to topological data analysis. I am particularly excited about the potential for collaboration in a postdoctoral environment, whether through joint work on theoretical questions (multiparameter persistence, stability theory), applied projects (neuroscience, complex systems), or software development (integration with existing TDA pipelines). My systematic approach to building foundations—exemplified by extending coordinates from dimension one to arbitrary dimensions—will enable me to tackle new challenges in applied topology while maintaining mathematical rigor. Ultimately, my research aims to make topological methods more accessible and interpretable for the broader scientific community. By providing explicit, computable geometric representations of abstract topological features, I hope to accelerate the adoption of TDA across diverse application domains and inspire new mathematical developments at the intersection of topology, geometry, and data science.

References

[1] Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. “Persistent Cohomology and Circular Coordinates”. In: *Discrete & Computational Geometry* 45.4 (2011), pp. 737–759. DOI: 10.1007/s00454-011-9344-x.

[2] Vin de Silva and Mikael Vejdemo-Johansson. “Persistent cohomology and circular coordinates”. In: *Proceedings of the twenty-fifth annual symposium on Computational geometry*. SoCG ’09. ACM, June 2009, pp. 227–236. DOI: 10.1145/1542362.1542406. URL: http://dx.doi.org/10.1145/1542362.1542406.

[3] Jose A. Perea. “Sparse Circular Coordinates via Principal $\mathbb{Z}$-Bundles”. In: *Topological Data Analysis*. Vol. 15. The Abel Symposium 1. Springer, 2020, pp. 435–458. DOI: 10.1007/978-3-030-43408-3_17.

[4] Jose A. Perea. “Multiscale Projective Coordinates via Persistent Cohomology of Sparse Filtrations”. In: *Discrete & Computational Geometry* 59.1 (2018), pp. 175–225. DOI: 10.1007/s00454-017-9927-2.

[5] Luis Polanco and Jose A. Perea. “Coordinatizing Data With Lens Spaces and Persistent Cohomology”. In: *Proceedings of the 31st Canadian Conference on Computational Geometry (CCCG 2019)*. arXiv preprint arXiv:1905.00350. 2019, pp. 49–57. URL: https://arxiv.org/abs/1905.00350.

[6] Luis Scoccola et al. “Toroidal Coordinates: Decorrelating Circular Coordinates With Lattice Reduction”. In: *39th International Symposium on Computational Geometry (SoCG 2023)*. Vol. 258. Leibniz International Proceedings in Informatics (LIPIcs). 2023, 57:1–57:20. DOI: 10.4230/LIPIcs.SoCG.2023.57.

[7] Frédéric Chazal et al. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer Cham, 2016. ISBN: 978-3-319-42543-6. DOI: 10.1007/978-3-319-42545-0. URL: https://link.springer.com/book/10.1007/978-3-319-42545-0.

[8] Jose A. Perea, Luis Scoccola, and Christopher J. Tralie. “DREiMac: Dimensionality Reduction with Eilenberg-MacLane Coordinates”. In: *Journal of Open Source Software* 8.91 (2023), p. 5791. DOI: 10.21105/joss.05791. URL: https://doi.org/10.21105/joss.05791.

[9] Ann E. Sizemore et al. “Cliques and cavities in the human connectome”. In: *Journal of Computational Neuroscience* 44.1 (Feb. 2018), pp. 115–145. DOI: 10.1007/s10827-017-0672-6.

[10] Giovanni Petri et al. “Homological scaffolds of brain functional networks”. In: *Journal of the Royal Society Interface* 11.20140873 (2014). DOI: 10.1098/rsif.2014.0873. URL: https://doi.org/10.1098/rsif.2014.0873.

[11] Sara Kališnik. “Tropical Coordinates on the Space of Persistence Barcodes”. In: *Foundations of Computational Mathematics* 19 (2019), pp. 101–129. DOI: 10.1007/s10208-018-9379-y.

[12] Steve Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. Vol. 209. Mathematical Surveys and Monographs. American Mathematical Society, 2015.

[13] Peter Bubenik and Jose A. Velez-Marulanda. “Exact weights and path metrics for triangulated categories and the derived category of persistence modules”. In: *Journal of Pure and Applied Algebra* 229.2 (2025), p. 107903. DOI: 10.1016/j.jpaa.2025.107903.

[14] Peter Bubenik and Jonathan A. Scott. “Categorification of persistent homology”. In: *Discrete & Computational Geometry* 51.3 (2014), pp. 600–627. DOI: 10.1007/s00454-014-9605-x.

[15] Justin Michael Curry. “Sheaves, Cosheaves and Applications”. arXiv:1303.3255. PhD thesis. University of Pennsylvania, 2014. URL: https://arxiv.org/abs/1303.3255.