Spectral families of persistent rank invariants

Matt Piekenbrock* and Jose A. Perea†

Abstract: *Using a duality result between persistence diagrams and persistence measures, we introduce a framework for constructing families of continuous relaxations of the persistent rank invariant for parametrized families of persistence vector spaces indexed over the real line. Like the rank invariant, these families obey inclusion-exclusion, are derived from simplicial boundary operators, and encode all the information needed to construct a persistence diagram. Unlike the rank invariant, these spectrally-derived families enjoy a number of stability and continuity properties typically reserved for persistence diagrams, such as smoothness and differentiability over the positive semi-definite cone. Leveraging a connection to combinatorial Laplacian operators, we find the non-harmonic spectra of our proposed relaxation encode valuable geometric information about the underlying space, prompting several avenues for geometric data analysis.*

Background: Persistent homology pipelines typically follow a well-established pattern: given an input data set X, construct a filtration (K, f) from X such that useful topological or geometric information may be profitably gleaned from its *persistence diagram*—a multiset summary of (K, f) constructed by pairing homological critical values $\{a_i\}_{i=1}^n$ with non-zero *multiplicities* $\mu_p^{i,j}$ or Betti numbers $\beta_p^{i,j}$ [5]:

$$\mathrm{dgm}_p(f) \triangleq \{\, (a_i, a_j) : \mu_p^{i,j} \neq 0 \,\} \,\cup\, \Delta$$
$$\mu_p^{i,j} \triangleq \left(\beta_p^{i,j\text{-}1} - \beta_p^{i,j}\right) - \left(\beta_p^{i\text{-}1,j\text{-}1} - \beta_p^{i\text{-}1,j}\right)$$

By pairing simplices using homomorphisms between homology groups, diagrams demarcate homological features succinctly. The essential quality of persistence is that this pairing exists, is unique, and is stable under additive perturbations [5]. Persistence is the de facto connection between homology and the application frontier.

Though theoretically sound, diagrams suffer from several practical issues: they are sensitive to strong outliers, far from injective, expensive to compute, and expensive to compare. Practitioners have tackled some of these issues by equipping diagrams with additional structure by way of maps to function spaces—examples include persistence landscapes [2] and references therein. These diagram vectorizations have proven useful for learning applications due to their stability and metric configurability. The scalability issue remains exacerbated though, as these vectorizations require diagrams as part of their input.

Approach: Rather than adding structure to precomputed diagrams, we propose a spectral method that performs both steps, simultaneously and approximately. Our approach constructs a vector-valued mappings over a *parameter space* $\mathcal{A} \subset \mathbb{R}^d$:

$$(X_\alpha, \mathcal{R}, \epsilon, \tau) \mapsto \mathbb{R}^{O(|\mathcal{R}|)}$$

where $\{X_\alpha\}_{\alpha\in\mathcal{A}}$ is an $\mathcal{A}$-parametrized family of data sets, $\mathcal{R} \subset \Delta_+$ a rectilinear *sieve* over the upper half-plane Δ_+, and $(\epsilon, \tau) \in \mathbb{R}_+^2$ are approximation/smoothness parameters, respectively. Our strategy is motivated by measure-theoretic perspectives on $\mathbb{R}$-indexed persistence modules [3, 4], which generalize $\mu_p^{i,j}$ to arbitrary *corner points* $(\hat{\imath}, \hat{\jmath}) \in \Delta_+$:

$$\mu_p^{\hat{\imath},\hat{\jmath}} = \min_{\delta>0} \left(\beta_p^{\hat{\imath}+\delta,\hat{\jmath}-\delta} - \beta_p^{\hat{\imath}+\delta,\hat{\jmath}+\delta}\right) - \left(\beta_p^{\hat{\imath}-\delta,\hat{\jmath}-\delta} - \beta_p^{\hat{\imath}-\delta,\hat{\jmath}+\delta}\right)$$

and also by a technical observation that shows the multiplicity function is expressible as a sum of *unfactored* boundary operators $\partial_p : C_p(K) \to C_{p-1}(K)$—that is, given a fixed $p \geq 0$, a filtration $K = \{K_i\}_{i\in[N]}$ of size $N = |K|$, and a rectangle $R = [i, j] \times [k, l] \subset \Delta_+$, the p-th multiplicity μ_p^R of K is given by:

$$\mu_p^R = \mathrm{rank} \begin{bmatrix} \partial_{p+1}^{j+1,k} & 0 \\ 0 & \partial_{p+1}^{i+1,l} \end{bmatrix} - \mathrm{rank} \begin{bmatrix} \partial_{p+1}^{i+1,k} & 0 \\ 0 & \partial_{p+1}^{j+1,l} \end{bmatrix}$$

where $\partial_p^{i,j}$ denotes the lower-left submatrix of ∂_p defined by the first j columns and the last $m-i+1$ rows. An explicit proof of this can be found in [6], though it was also noted in passing by Edelsbrunner [7]—it can proved by combining the Pairing Uniqueness Lemma with the fact that left-to-right column operations preserves the ranks of “lower-left” submatrices. Though often used to show the correctness of the reduction algorithm from [7], the implications of this fact are quite general, as noted recently by Bauer et al. [1]:

Proposition 1 ([1]). *Any persistence algorithm which preserves the ranks of the submatrices $\partial^{i,j}(K_\bullet)$ for all $i, j \in [N]$ is a valid persistence algorithm.*

*Khoury College of Computer Sciences, Northeastern University.

†Department of Mathematics and Khoury College of Computer Sciences, Northeastern University.

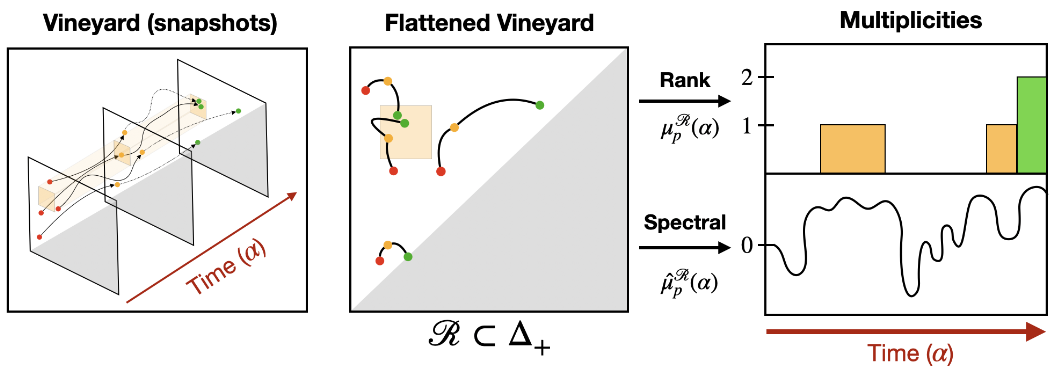

Figure 1: (Left) Vineyards analogy of diagrams at 'snapshots' over time; (middle) vineyard curves flattened with a sieve $\mathcal{R} \subset \Delta_+$; (right) the integer-valued multiplicity function $\mu_p^{\mathcal{R}}(f_\alpha)$ as a function of time $\alpha \in \mathbb{R}$ (top) and a real-valued spectral relaxation (bottom)

Spectral rank invariant: Our proposed mapping exploits proposition 1 via a spectral characterization of μ_p^R. In particular, let K denote a fixed simplicial complex constructed from a data set X and f_α a continuous filter function satisfying, for all $\alpha \in \mathcal{A}$:

$$(K, f_\alpha) \triangleq \{ f_\alpha : K \to \mathbb{R} \mid f_\alpha(\tau) \leq f_\alpha(\sigma), \tau \subseteq \sigma \in K \}$$

Our methods inputs are (K, f_α), a sieve $\mathcal{R} \subset \Delta_+$, and parameters $(\epsilon, \tau) \in \mathbb{R}_+^2$ representing how *closely* and *smoothly* the relaxation should model the quantity:

$$\mu_p^{\mathcal{R}}(f_\alpha) \triangleq \left\{ \text{card}\left(\text{dgm}_p(f_\alpha)\big|_{\mathcal{R}} \right) \mid \alpha \in \mathcal{A} \right\}$$

The intuition is that $\mathcal{R}$ filters and summarizes topological and geometric behavior exhibited by X_α for all $\alpha \in \mathcal{A}$, thereby *sifting* the space $\mathcal{A} \times \Delta_+$. Our proposed approximation first associates a *normalized combinatorial Laplacian* operator $\mathcal{L} : C^p(K, \mathbb{R}) \to C^p(K, \mathbb{R})$ to the corner points on the boundary of $\mathcal{R}$. Then, for some $v \in \text{span}(\mathbf{1})^\perp$, we restrict and project $\mathcal{L}$ onto the following Krylov subspace:

$$\mathcal{K}_n(\mathcal{L}, v) \triangleq \text{span}\{v, \mathcal{L}v, \mathcal{L}^2 v, \ldots, \mathcal{L}^{n-1} v\}$$

We can show (1) the eigenvalues of $T = \text{proj}_{\mathcal{K}} \mathcal{L}\big|_{\mathcal{K}}$ provide an $(1-\epsilon)$-approximation of $\mu_p^{\mathcal{R}}(f_\alpha)$, and (2) varying $\tau > 0$ yields a family of spectral operators whose Schatten-1 norms are Lipshitz continuous, stable under relative perturbations, and differentiable on the positive semi-definite cone. Moreover, as the parameters ϵ and τ approach zero, the multiplicity $\mu_p^{i,j}$ is recovered exactly.

Unlike existing dynamic persistence algorithms, our approach requires no complicated data structures or maintenance procedures to implement, can be made *matrix-free*, and is particularly efficient to compute over parameterized families of inputs. We defer the formal analysis, properties, and applications of the method to full[1] paper, in preparation.

References

[1] Ulrich Bauer, Talha Bin Masood, Barbara Giunti, Guillaume Houry, Michael Kerber, and Abhishek Rathod. Keeping it sparse: Computing persistent homology revised. *arXiv preprint arXiv:2211.09075*, 2022.

[2] Peter Bubenik et al. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015.

[3] Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi. Betti numbers in multidimensional persistent homology are stable functions. *Mathematical Methods in the Applied Sciences*, 36(12):1543–1557, 2013.

[4] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*, volume 10. Springer, 2016.

[5] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

[6] Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis.* Cambridge University Press, 2022.

[7] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE, 2000.

[1]The paper in preparation can be found at: `https://github.com/peekxc/pbsig/blob/main/notes/pbsig.pdf`