1 Introduction

Motivation:

Organization: In what follows, we investigate how to formulate the persistent Betti number (PBN) computation on filtered simplicial complexes in parameterized settings. In particular, in section 3, we show how the PBN may be expressed as a sum of rank computations on unfactored boundary (sub)matrices. By representing the boundary operators implicitly and exploiting a variety of properties of the rank function, we demonstrate the rank-based PBN expression admits certain computational advantages in both static and parameterized settings. We show the p-th PBN of a filtered simplicial complex $K_\bullet$ may be computed in essentially quadratic time and linear storage, as opposed to the cubic time and quadratic storage complexity required by persistence. Finally, we illustrate both the computational and practical advantages our expression has with a few applications in section 6.

2 Background & Notation

A *simplicial complex* $K \subseteq \mathcal{P}(V)$ over a vertex set $V = \{v_1, v_2, \ldots, v_n\}$ is a collection of simplices $\{\sigma : \sigma \in \mathcal{P}(V)\}$ such that $\tau \subseteq \sigma \in K \implies \tau \in K$. We denote with $K^{(p)} = \{\, \sigma \in K : \dim(\sigma) = p \,\}$ the p-simplices of K. A *filtration* $K_\bullet = \{K_i\}_{i \in I}$ of a simplicial complexes indexed by a totally ordered set I is a family of complexes such that $i < j \in I \implies K_i \subseteq K_j$. $K_\bullet$ is called *simplexwise* if $K_j \smallsetminus K_i = \{\sigma_j\}$ whenever j is the immediate successor of i in I and $K_\bullet$ is called *essential* if $i \neq j$ implies $K_i \neq K_j$:

$$\emptyset = K_0 \subsetneq K_1 \subsetneq \cdots \subsetneq K_m = K_\bullet, \quad K_i = K_{i-1} \cup \{\sigma_i\} \tag{1}$$

Filtrations may be equivalently defined via *filter functions* $f : K \to I$ satisfying $f(\tau) \leq f(\sigma)$ whenever $\tau \subseteq \sigma$. Here, we consider two index sets for I: $\mathbb{R}$ and $[n] = \{1, \ldots, n\}$. Any finite filtration may be trivially converted into an essential, simplexwise filtration via a set of *condensing*, *refining*, and *reindexing* maps [1]. For simplicity, but without loss of generality, we exclusively consider essential simplexwise filtrations and for brevity refer to them as filtrations.

For K a simplicial complex and $\mathbb{F}$ a field, a p-chain is a formal $\mathbb{F}$-linear combination of p-simplices of K. The collection of p-chains under addition yields an $\mathbb{F}$-vector space denoted $C_p(K)$. The p-boundary $\partial_p(\sigma)$ of an oriented p-simplex $\sigma \in K$ is defined as the alternating sum of its oriented co-dimension 1 faces:

$$\partial_p(\sigma) = \partial_p([v_0, v_1, \ldots, v_p]) := \sum_{i=0}^{p} (-1)^i [v_0, \ldots, v_{i-1}, v_{i+1}, \ldots v_p] \tag{2}$$

The p-boundary of a p-chain is defined linearly in terms of its constitutive simplices. A p-chain c with zero boundary $\partial_p c = 0$ is called a p-cycle. Together, the collection of p-boundaries and p-cycles forms the groups $B_p(K) = \operatorname{Im} \partial_{p+1}$ and $Z_p(K) = \operatorname{Ker} \partial_p$, respectively. Since $\partial_p \circ \partial_{p+1} = 0$ for all $p \geq 0$, the quotient space $H_p(K) = Z_p(K)/B_p(K)$ is well-defined, and $H_p(K)$ is called the p-th homology of K with coefficients in $\mathbb{F}$. The dimension of the p-th homology group $\beta_p(K) = \dim(H_p(K))$ of K is called the p-th *Betti* number of K.

Let $K_\bullet = \{K_i\}_{i \in [m]}$ denote a filtration of size $|K_\bullet| = m$, and let $\Delta_+^m = \{(i, j) : 0 \leq i \leq j \leq m\}$ denote the set of filtration index pairs. For every such pair $(i, j) \in \Delta_+^m$, the inclusions $K_i \subsetneq K_{i+1} \subsetneq \cdots \subsetneq K_j$ induce linear transformations $h_p^{i,j}$ at the level of homology:

$$0 = H_p(K_0) \to \cdots \to H_p(K_i) \underbrace{\to \cdots \to}_{h_p^{i,j}} H_p(K_j) \to \cdots \to H_p(K_m) = H_p(K_\bullet) \tag{3}$$

When $\mathbb{F}$ is a field, this sequence of homology groups uniquely decomposes $K_\bullet$ into a pairing of simplices (σ_i, σ_j) demarcating the evolution of homology classes [15]: σ_i marks the creation of a homology class, σ_j marks its destruction, and the difference $|i - j|$ records the lifetime of the class, called its *persistence*. The p-th persistent homology groups are the images of these transformations and the p-th persistent Betti numbers are their dimensions:

$$H_p^{i,j} = \begin{cases} H(K_i) & i = j \\ \operatorname{Im} h_p^{i,j} & i < j \end{cases}, \qquad \beta_p^{i,j} = \begin{cases} \beta_p(K_i) & i = j \\ \dim(H_p^{i,j}) & i < j \end{cases} \tag{4}$$

For a fixed $p \geq 0$, the collection of persistent pairs (i, j) together with unpaired simplices (l, ∞) form a summary representation $\operatorname{dgm}_p(K_\bullet)$ called the *p-th persistence diagram of $K_\bullet$*. Conceptually, $\beta_p^{i,j}$ counts the number of persistent pairs lying inside the box $(-\infty, i\,] \times (\, j, \infty)$ (see Figure 1)—the number of persistent homology groups born at or before i that died sometime after j.

The duality between PBNs and Diagrams

The connection between the persistent homology (PH) groups and their corresponding persistent Betti numbers (PBNs) has long been studied from multiple perspectives by several authors [2, 3, 6, 15]. From an algebraic perspective, Carlsson et al. [15] observed that the PH groups over a filtration may be viewed as the standard homology groups of a particular graded module M over a polynomial ring. In [6], Cohen-Steiner et al. give a more discrete perspective on PH by defining the persistence diagram in terms of a *multiplicities*: given a tame function $f : \mathcal{X} \to \mathbb{R}$ over a topological space $\mathcal{X}$, its homological critical values $\{a_i\}_{i=1}^n$, and an interleaved sequence $\{b_i\}_{i=0}^n$ satisfying $b_{i-1} < a_i < b_i$ for all $1 \leq i \leq n$, the p-th persistence diagram over f is given as:

$$\mathrm{dgm}_p(f) = \{\, (a_i, a_j) : \mu_p^{i,j} \neq 0 \,\} \cup \Delta \tag{5}$$

where Δ denotes the diagonal, counted with infinite multiplicity, and $\mu_p^{i,j}$ is the *multiplicity function*, defined as:

$$\mu_p^{i,j} = \left(\beta_p^{i,j-1} - \beta_p^{i,j}\right) - \left(\beta_p^{i-1,j-1} - \beta_p^{i-1,j}\right) \qquad \text{for } 0 \leq i < j \leq n+1 \tag{6}$$

Equation (6) illuminates an intrinsic connection between the multiplicity function and the persistent Betti numbers. Namely, the inclusion-exclusion property that (6) obeys suggests that diagrams completely characterize their PBNs. Indeed, the fundamental lemma of persistent homology [9] states that for every pair of indices $0 \leq k \leq l \leq n+1$:

$$\beta_p^{k,l} = \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j} \tag{7}$$

Thus, if one is interested in computing any of the PBNs of some space $\mathcal{X}$, it is sufficient to compute $\mathrm{dgm}_p(\mathcal{X})$ and read them off directly via (7). Perhaps as a result of this characterization, much of persistence-related research has concentrated on exploiting properties of the diagram itself [], as opposed to the PBNs ([2, 4] are notable exceptions). Nonetheless, as we shall show, there are certain advantages the PBN computation has over the "standard" reduction algorithm from [9]. Indeed, (6) implies one in theory recover the diagram through [a finite number of] PBN computations alone, suggesting an alternative computational paradigm with which to approach persistence computation [4].

For persistence modules indexed over the real line, the duality between diagrams and PBNs was further studied by Chazal [3] in the measure-theoretic setting. By reinterpreting the multiplicity function μ_p^* as a certain kind of integer-valued measure over rectangles in the plane, a generalization of (6) was shown by demonstrating that one may recover the diagram of a persistence module M over $\mathbb{R}$ by constructing its corresponding *persistence measure*:

$$\mu_p(R; M) = \mathrm{card}\left(\mathrm{dgm}_p(M)\big|_R\right) \qquad \text{for all rectangles } R \subset \mathbb{R}^2 \tag{8}$$

Cerri et al. [2] incorporate this interpretation in their work studying the stability of PBNs in multidimensional persistence by showing that *proper cornerpoints* in the persistence diagram are points $x = (i, j) \in \Delta_+$ satisfying:

$$x = (i,j) \in \mathrm{dgm}_p(f) \iff \mu_p(x) > 0 \iff \min_{\epsilon > 0} \left(\beta_p^{i+\epsilon, j-\epsilon} - \beta_p^{i+\epsilon, j+\epsilon}\right) - \left(\beta_p^{i-\epsilon, j-\epsilon} - \beta_p^{i-\epsilon, j+\epsilon}\right) > 0 \tag{9}$$

One may compare (6) with (9). One of the primary contributions from [2] is a representation theorem akin to (7) (Theorem 3.11) expressing the persistent Betti number function $\beta_* : \Delta_+ \to \mathbb{N} \cup \{\infty\}$ as a sum of multiplicity functions. A consequence of this theorem is that any distance between persistence diagrams induces a distance between PBN functions—if X is a triangulable space and $f, g : X \to \mathbb{R}$ are two continuous functions, then $d(\beta_f, \beta_g) \leq \max_{x \in X} |f(x) - g(x)|$, where:

$$d(\beta_f, \beta_g) = \inf_{\phi} \sup_{p \in \mathrm{dgm}(f)} \|p - \phi(p)\|_\infty$$

is the (extended) matching distance between PBN functions (β_f, β_g), ϕ ranges over all multi-bijections between $\mathrm{dgm}(f)$ and $\mathrm{dgm}(g)$, and $\|\cdot\|_\infty$ measures the pseudo-distance [2] between points. Thus, PBN functions are stable functions: small changes in continuous scalar-valued filtering functions imply small changes in the corresponding persistent Betti numbers functions. Thus, like persistence diagrams, this stability justifies their use and study of PBNs in continuously parameterized settings.

3 Motivating Derivation

Let $B_p(K_\bullet) \subseteq Z_p(K_\bullet) \subseteq C_p(K_\bullet)$ denote the p-th boundary, cycle, and chain groups of a given filtration $K_\bullet$, respectively. Additionally, let $\partial_p : C_p(K_\bullet) \to C_p(K_\bullet)$ denote the boundary operator sending p-chains to their

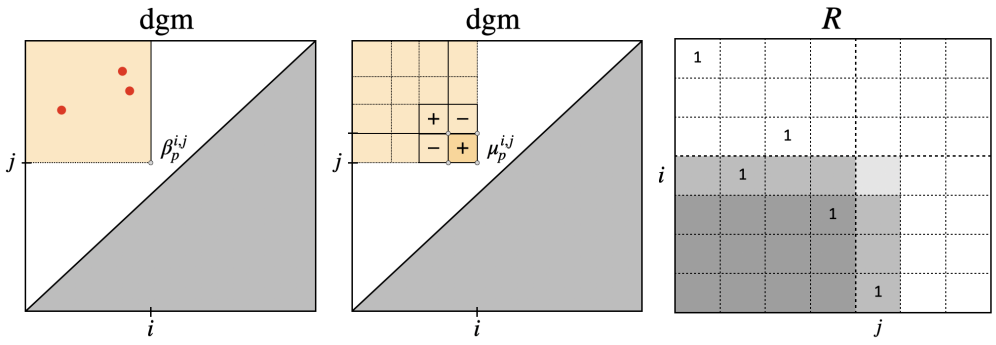

Figure 1: (Left) the persistent Betti number $\beta_p^{i,j}$ counts the number of points (3) in upper left-corner of $\mathrm{dgm}_p(K_\bullet)$. (Middle) The additivity of PBNs can be used to express multiplicity $\mu_p^{i,j}$ of any given box. (Right) The computational interpretation of the Pairing Uniqueness Lemma; in this case $r_R(i,j) = 3 - 2 + 1 - 2 = 0$ yields whether the entry $R[i,j]$ is non-zero.

respective boundaries. With a slight abuse of notation, we also use ∂_p to also denote the filtration boundary matrix with respect to an ordered basis $(\sigma_i)_{1 \le i \le m_p}$. The p-th persistent Betti number $\beta_p^{i,j}$ at index $(i,j) \in \Delta_+^m$, where $\Delta_+^m := \{\, (i,j) \in [m] \times [m] : i < j \,\}$, is defined as:

$$\begin{aligned} \beta_p^{i,j} = \dim(H_p^{i,j}) &= \dim\left(Z_p(K_i)/B_p(K_j)\right) \\ &= \dim\left(Z_p(K_i)/(Z_p(K_i) \cap B_p(K_j)\right) \\ &= \dim\left(Z_p(K_i)\right) - \dim\left(Z_p(K_i) \cap B_p(K_j)\right) \end{aligned} \tag{10}$$

While $Z_p(K_i) = \mathrm{nullity}(\partial_p(K_i))$ and thus reduces to a matrix rank computation, the intersection term (the persistence part) is more subtle. Zomorodian et al. [15] outline an algorithm to compute a basis for $Z_p(K_i) \cap B_p(K_j)$ via a sequence of boundary matrix reductions; the subsequent Theorem (5.1) reduces the complexity of computing PH groups with coefficients in any PID to that of computing homology groups. The approach is essentially the same as the subspace intersection algorithm presented in []. However, the standard homology computations require $O(m^2)$ space and $O(m^3)$ time to compute, implying either of these approaches to computing the PBN computation exhibits same complexity as the full persistence computation.

In what follows, we outline a different approach to computing (10) that is both simpler and computationally more attractive. To illustrate our approach, we require more notation. If A is a $m \times n$ matrix, let $A^{i,j}$ denote the lower-left submatrix defined by last $m - i + 1$ rows (rows i through m, inclusive) and the first j columns. For any $1 \le i < j \le m$, define the quantity $r_A(i,j)$ as follows:

$$r_A(i,j) = \mathrm{rank}(A^{i,j}) - \mathrm{rank}(A^{i+1,j}) + \mathrm{rank}(A^{i+1,j-1}) - \mathrm{rank}(A^{i,j-1}) \tag{11}$$

The structure theorem from [15] shows that 1-parameter persistence modules can be decomposed in an *essentially unique* way into indecomposables. Computationally, a consequence of this phenomenon is the Pairing Uniqueness Lemma [7], which asserts that if $R = \partial V$ is the decomposition of the boundary matrix, then:

$$r_R(i,j) \neq 0 \Leftrightarrow R[i,j] \neq 0$$

Since the persistence diagram is derived completely from R, this suggests that information about a diagram can be obtained through rank computations alone. For a more geometric description of this idea, see the third picture in Figure 1. We record a non-trivial fact that follows from this observation:

Lemma 1 (Dey & Wang [8])**.** *Let $R = \partial V$ denote the matrix decomposition of a given filtered boundary matrix ∂ derived from the associated filtration $K_\bullet$. For any pair (i,j) satisfying $1 \le i < j \le m$, we have:*

$$\mathrm{rank}(R^{i,j}) = \mathrm{rank}(\partial^{i,j}) \tag{12}$$

Equivalently, all lower-left submatrices of ∂ have the same rank as their corresponding submatrices in R.

Lemma 1 was the essential motivating step used by Chen et al [4] in their rank-based persistence algorithm—the first output-sensitive algorithm given for computing persistent homology of a filtered complex. Though (12) is defined

over the full boundary matrix ∂, there is no loss in generality in extending this lemma to p-dimensional homology, for any fixed choice $p \geq 0$. To see this, note that since the reduction algorithm only adds p-chains to p-chains, we may simply ignore all q-chains for $q \neq p$. That is, if the p-chain of a p-simplex σ_j corresponds to column j in ∂, the only columns that may be added to j must correspond to simplices of dimension p. Hence, if we set all columns corresponding to simplices of dimension $q \neq p$ to 0 in the $m \times m$ boundary matrix ∂, then ∂ represents the p-th boundary operator $\partial_p : C_p(K_\bullet) \to C_{p-1}(K_\bullet)$.

In what follows, we will use ∂_p and R_p to refer to matrices of ∂ and R whose q-chains are set to 0, for $q \neq p$. The first fact we prove is that Lemma 1 may be used to write the persistent Betti number as a sum of rank functions.

Proposition 1. *Given a fixed $p \geq 0$, a filtration $K_\bullet$ of size $m = |K_\bullet|$, and any pair $(i, j) \in \Delta_+^m$, the persistent Betti number $\beta_p^{i,j}(K_\bullet)$ at (i, j) is given by:*

$$\beta_p^{i,j}(K_\bullet) = i - \operatorname{rank}(\partial_p^{1,i}) - \operatorname{rank}(\partial_{p+1}^{1,j}) + \operatorname{rank}(\partial_{p+1}^{i+1,j}) \tag{13}$$

A detailed proof of Proposition 1 is given in the appendix. It turns out Lemma 1 can also be used to generalize (13) to arbitrary rectangles in Δ_+ via *μ-queries*: box-parameterized rank queries which count the number of persistence pairs that intersect a fixed "box" placed in the upper half-plane [4]. As a result, the p-th multiplicity function can also be defined in a rank-based formulation akin to (13).

Proposition 2 (Chen & Kerber [4])**.** *Given a fixed $p \geq 0$, a filtration $K_\bullet$ of size $m = |K|$, and a $R = [i, j] \times [k, l]$ whose indices (i, j, k, l) satisfy $0 \leq i < j \leq k < l \leq m$, the p-th multiplicity μ_p^R of $K_\bullet$ is given by:*

$$\mu^R(K_\bullet) = \operatorname{rank}(\partial_{p+1}^{j+1,k}) - \operatorname{rank}(\partial_{p+1}^{i+1,k}) - \operatorname{rank}(\partial_{p+1}^{j+1,l}) + \operatorname{rank}(\partial_{p+1}^{i+1,l}) \tag{14}$$

Compared to other methods of obtaining $\beta_p^*(K_\bullet)$ and $\mu_p^*(K_\bullet)$ (such as those in [9, 15]), the primary advantage the rank-based expressions from (13)-(14) have is that their computations are performed directly on *unfactored* boundary matrices. Thus, we may exploit various properties of the rank function to accelerate (or relax) the computation of both invariants without e.g. reducing ∂ to R, zeroing columns of ∂ to obtain ∂_p, or even representing ∂ in memory explicitly. We dedicate the rest of the paper to exploring the consequences of this fact.

Boundary matrices are sparse and highly structured: given $K_\bullet$ with m simplices $\sigma_1, \sigma_2, \ldots, \sigma_m$ constructed from a p-dimensional complex K, its full boundary matrix ∂ is upper-triangular and has a storage complexity of:

$$\operatorname{nnz}(\partial) \sim O(m \log m) \tag{15}$$

To see this, note that since a p-simplex has $2^{p+1} - 1$ faces, we have $p \leq \log(m + 1) - 1$. Moreover, since each column of ∂_p^* contains exactly $p + 1$ non-zero entries, $\operatorname{nnz}(\partial) \sim O((p + 1)m)$. As a result, we have the following corollary:

Corollary 1. *Given a filtration $K_\bullet$ of size $m = |K_\bullet|$ and indices $i, j \in \Delta_+^m$, computing $\beta_p^{i,j}$ using expression (13) requires $O(\mathrm{R}_p(j))$ time, where $\mathrm{R}_p(k)$ denotes the complexity of computing the rank of square $k \times k$ matrix with $O((p + 1)k)$ non-zero $\mathbb{F}$ entries.*

The relation $\partial_{p+1}^{i+1,j} \subseteq \partial_{p+1}^{1,j}$ implies the largest number of non-zeros in any matrix in (13) is $O(\max\{p\, i, (p + 1)\, j\})$.

Corollary 2. *Given a filtration $K_\bullet$ of size $m = |K_\bullet|$ and a rectangle $R = [i, j] \times [k, l]$ with indices $0 \leq i < j \leq k < l \leq m$, computing μ_p^R using expression (14) requires $O(\mathrm{R}_p(l - i))$ time, where $\mathrm{R}_p(k)$ denotes the complexity of computing the rank of square $k \times k$ matrix with $O((p + 1)k)$ non-zero $\mathbb{F}$ entries.*

In the next section, we show how extend these complexity statements to parameterized settings. We address the storage and time complexity of $R_\partial(\cdot, \cdot)$ in section 2.

A Parameterized Boundary Matrix Relaxation: Expressing the PBN via (13) enables us to exploit properties of the rank function which are advantageous in *parameterized* settings, i.e. settings where the input data is thought to be generated from a parameterized family. One such property is permutation invariance: given any $A \in \mathbb{R}^{n \times n}$, it is well known that $\operatorname{rank}(A) = \operatorname{rank}(P^T A P)$ for any permutation matrix P. Though the boundary matrices ∂_p in (13) are given in filtration order to elucidate their structure, the permutation invariance of the rank function suggests they need not be in such an order to be evaluated—so long as the constitutive terms have the same non-zero pattern as their filtration-ordered counterparts, their ranks and thus their PBNs will be identical. In what follows, we re-define the boundary matrix to exploit this permutation invariance.

All of the notation given thus far, such as the definitions of the PH groups (3) and the the PBN (10), have used integer indices $(i, j) \in \Delta_+^m$ to describe the PH groups over a filtration pair $(K_\bullet, f)$ of size $|K| = m$. Equivalently, we

have thus far implicitly assumed the range of the filter function $f : K \to I$ to be the typical index set $I = [m]$. In practice, the filter function f is often derived from geometrical settings wherein it is more informative to interpret the persistence of a persistent-pair $(\sigma_i, \sigma_j) \in \mathrm{dgm}(K_\bullet)$ as $f(\sigma_j) - f(\tau_i)$, rather than as $j - i$, and it is more pragmatic to think of $f : K \to \mathbb{R}$ is real-valued. In theory, any function satisfying $f(\tau) \subseteq f(\sigma)$ for every face/coface pair (τ, σ) yields a well defined persistence diagram. In what follows, in the spirit of (9), we alter our notation by re-defining $\beta_p^{i,j}$ using pairs $(i, j) \in \Delta_+$ from the upper-half plane $\Delta_+ = \{(x, y) \in \mathbb{R}^2 : y > x\}$.

Suppose that instead of being given a fixed pair $(K_\bullet, f)$, the filter function was parameterized $f : \mathcal{H} \times K \to \mathbb{R}$ and one was interested in computing $\beta_p^{i,j}$ over $\mathcal{H}$. We give several application contexts where this kind of formulation occurs naturally in section 6. As a first step to simplifying the PBN computation in this setting, we introduce the notion of a *parameterized boundary matrix*.

Definition 1 (Parameterized boundary matrix)**.** *Let K denote an abstract simplicial complex of size $|K| = m$, equipped with parameterized filtering function $f : K \times \mathcal{H} \to \mathbb{R}$. Assume K is ordered along a fixed but arbitrary linear extension $(K, \preceq^*)$ of the face poset of K. For fixed $(i, j) \in \Delta_+$, define the* $\mathcal{H}$-parameterized p-th boundary matrix $\hat{\partial}_p^{i,j}(h)$ at scale (i, j) *to be the $m \times m$ matrix ordered by $\preceq^*$ for all $h \in \mathcal{H}$, and whose entries (k, l) satisfy:*

$$\hat{\partial}_p^{i,j}(h)[k,l] = \begin{cases} \pm\,(S_i \circ f_h)(\sigma_k)) \cdot (\tilde{S}_j \circ f_h)(\sigma_l) & \text{if } \sigma_k \in \partial_p(\sigma_l) \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $S_i : \mathbb{R} \to \{0, 1\}$ is a step *function satisfying $S_i(x) = 0$ if $x \leq i$ and 1 otherwise, $\tilde{S}_i = 1 - S_i$, and $f_h(\sigma) = f(\sigma, h)$.*

The result of definition 1 is that, in parameterized settings, the PBN can be written in essentially the same form as (13). To simplify the notation, we write $A^x = A^{*,x}$ for the setting where only columns up to x of A are being selected, and let $q = p + 1$. We also write $r(A) = \mathrm{rank}(A)$. The parameterized PBN can be written as:

$$\begin{aligned} &\beta_p^{i,j} : \mathcal{H} \to \mathbb{N} \\ &\qquad h \mapsto |K_i^{(p)}(h)| - (\mathrm{r} \circ \hat{\partial}_p^i)(h) - (\mathrm{r} \circ \hat{\partial}_q^j)(h) + (\mathrm{r} \circ \hat{\partial}_q^{i+\epsilon,j})(h) \end{aligned} \tag{17}$$

where $\epsilon > 0$ is an arbitrarily small positive number. By the reduction shown above, we also have a parameterized multiplicity function for any rectangle $R = [i, j] \times [k, l]$ in the upper half-plane Δ_+ satisfying $i < j \leq k < l$:

$$\begin{aligned} &\mu_p^R : \mathcal{H} \to \mathbb{N} \\ &\qquad h \mapsto (\mathrm{r} \circ \hat{\partial}_q^{j+\epsilon,k})(h) - (\mathrm{r} \circ \hat{\partial}_q^{i+\epsilon,k})(h) - (\mathrm{r} \circ \hat{\partial}_q^{j+\epsilon,l})(h) + (\mathrm{r} \circ \hat{\partial}_q^{i+\epsilon,l})(h) \end{aligned} \tag{18}$$

Special Cases: The spectrum of the graph Laplacian is known to have closed-form expressions for certain structured graphs. In particular, the spectra of the cycle graph C_n and the path graph P_n over n vertices is given by:

$$\Lambda(C_n) = 1 - \cos\left(\frac{2\pi k}{n}\right), \quad \Lambda(P_n) = 1 - \cos\left(\frac{\pi k}{n-1}\right) \tag{19}$$

for all $k = 0, 1, \ldots, n-1$. Thus, if we know ahead of times the graph we are working with is one of these special graphs, it follows that may read off its (i, j)-th PBN in $O(n)$ time.

Complexity of Persistence:

An quadratic-time rank computation

In this section we discuss the computational details involved in evaluating the rank function on the boundary matrices whose columns take the form (2). For any $X \in \mathbb{R}^{n \times m}$ with $n < m$, the *rank* function is defined as:

$$\mathrm{rank}(X) = \sum_{i=1}^{n} \mathrm{sgn}_+(\sigma_i(X)), \qquad \mathrm{sgn}_+(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $\Sigma = \mathrm{diag}(\{\sigma_1, \sigma_2, \ldots, \sigma_n\})$ are the singular values from the singular value decomposition (SVD) $X = U\Sigma V^T$ and $\mathrm{sgn}_+ : \mathbb{R}_+ \to \{0, 1\}$ is the one-sided sign function. As (20) is defined completely in terms of the singular values of X, and the singular values of X are given by the square roots of eigenvalues of XX^T (or X^TX), we focus on iterative methods for finding the eigenvalues of a Hermitian matrices. In particular, we focus on the Lanczos iteration, and related adaptations to positive semi-definite, diagonally dominant (DD) and strictly diagonally dominant (SDD) cases. We also show how these methods made be adapted to efficiently compute (17) and (18).

The Lanczos iteration: For a real, square matrix A of order n, the quadratic form $x^T Ax$ defines a continuous real-valued function of $x \in \mathbb{R}^n$. When A is symmetric positive definite, the implicit equation $x^T Ax = 1$ defines an n-dimensional ellipsoid $y^T \Lambda y = 1$ whose n principle axes are eigenvectors $\{v_i\}_{i=1}^n$ and whose lengths are the squares of eigenvalues $\Lambda(A)$ of A. Each eigen-pair of (λ, v) satisfies $Av = \lambda v$, and when A is symmetric, every λ is real-valued and every pair of eigenvectors $v, u \in \mathbb{R}^n$ whose corresponding eigenvalues $\lambda \neq \lambda'$ are orthogonal. Thus we may reveal the spectrum $\Lambda(A)$ of A—effectively the lengths of the aforementioned ellipsoid—via orthogonal diagonalization:

$$A = V \Lambda V^T = \sum_{i=1}^{n} \lambda_i v_i v_i^T \tag{21}$$

Decomposing A as in (21) is known as the *symmetric eigenvalue problem*. Computing the eigen decompositions of symmetric matrices generally consist of two phases: (1) reduction to tridiagonal form $Q^T AQ = T$ via orthogonal similarity transformations $Q = Q_1 Q_2 \dots Q_{n-2}$, and (2) diagonalization of the tridiagonal form $T = Y \Theta Y^T$. The latter may be performed in $O(n \log n)$ time [11], whereas the former phase is effectively bounded below by $\Omega(n^3)$ for dense matrices using non-Strassen-like computations, and thus it is this reduction to tridiagonal form that dominates the computation. To counter this when A sparse and structured, Lanczos [] proposed the so-called *method of minimized iterations* to obtain a tridiagonal form of A in much lower complexity.

The means by which the Lanczos method estimates eigenvalues is by projecting onto successive Krylov subspaces. Given a large, sparse, symmetric $n \times n$ matrix A with eigenvalues $\lambda_1 \geq \lambda_2 > \cdots \geq \lambda_r > 0$ and a vector $v \neq 0$, the order-j Krylov subspaces of the pair (A, v) are the spaces spanned by:

$$\mathcal{K}_j(A, v) := \text{span}\{v, Av, A^2 v, \dots, A^{j-1}\} = \text{range}(K_j(A, v)) \tag{22}$$

where $K_j(A, v) = [v \mid Av \mid A^2 v \mid \cdots \mid A^{j-1}]$ are their corresponding Krylov matrices. Krylov subspaces arise naturally from using the minimal polynomial of A to express A^{-1} in terms of powers of A. In particular, if A is nonsingular and its minimal polynomial has degree m, then $A^{-1} v \in K_m(A, v)$ and $K_m(A, v)$ is an invariant subspace of A. Thus, since A is orthogonally diagonalizable, we may obtain $\Lambda(A)$ by generating an orthonormal basis for $\mathcal{K}_n(A, v)$. To do this, the Lanczos method constructs successive QR factorizations of $K_j(A, v) = Q_j R_j$ for each $j = 1, 2, \dots, n$. Due to A's symmetry and the orthogonality of Q_j, we have $q_k^T A q_l = q_l^T A^T q_k = 0$ for $k > l + 1$, implying the corresponding $T_j = Q_j^T A Q_j$ have a tridiagonal structure:

$$T_j = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{j-1} \\ & & & \beta_{j-1} & \alpha_j \end{bmatrix}, \ \beta_j > 0, \ j = 1, 2, \dots, n \tag{23}$$

Unlike the spectral decomposition $A = V \Lambda V^T$, which identifies a diagonalizable A with its spectrum $\Lambda(A)$ up to a change of basis $A \mapsto M^{-1} A M$, there is no canonical choice of T_j due to the arbitrary choice of v. However, there is a connection between the iterates $K_j(A, v)$ and the full tridiagonalization of A: if $Q^T AQ = T$ is tridiagonal and $Q = [\, q_1 \mid q_2 \mid \cdots \mid q_n \,]$ is an $n \times n$ orthogonal matrix $QQ^T = I_n$, then:

$$K_n(A, q_1) = QQ^T K_n(A, q_1) = Q[\, e_1 \mid Te_1 \mid T^2 e_1 \mid \cdots \mid T^{n-1} e_1 \,] \tag{24}$$

is the QR factorization of $K_n(A, q_1)$. Indeed, the Implicit Q Theorem [10] asserts that if an upper Hessenburg matrix $T \in \mathbb{R}^{n \times n}$ has only positive elements on its first subdiagonal and there exists an orthogonal matrix Q such that $Q^T AQ = T$, then Q and T are *uniquely* determined by (A, q_1). Thus, given an initial pair (A, q_1) satisfying $\|q_1\| = 1$, we may restrict and project A to its j-th Krylov subspace T_j via:

$$AQ_j = Q_j T_j + \beta_j q_{j+1} e_j^T \tag{25}$$

where $Q_j = [\, q_1 \mid q_2 \mid \cdots \mid q_j \,]$ is an orthonormal set of vectors mutually orthogonal to q_{j+1}, e_j is the standard j identity vector, and $\beta_j > 0$. Equating the j-th columns on each side of (25) and rearranging the terms yields the *three-term recurrence*:

$$\beta_j \, q_{j+1} = Aq_j - \alpha_j \, q_j - \beta_{j\text{-}1} \, q_{j\text{-}1} \tag{26}$$

where $\alpha_j = q_j^T A q_j$, $\beta_j = \|r_j\|_2$, $r_j = (A - \alpha_j I) q_j - \beta_{j\text{-}1} q_j$, and $q_{j+1} = r_j / \beta_j$. Equation (26) is a variable-coefficient second-order linear difference equation, and it is a known fact that such equations have unique solutions: if $(q_{j\text{-}1}, \beta_j, q_j)$

are known, then $(\alpha_j,\ \beta_{j+1}, q_{j+1})$ are completely determined. The sequential process that iteratively builds T_j by exploiting the recurrence from (26) is called the *Lanczos iteration*. Note that if A is singular and we encounter $\beta_j = 0$ for some $j < n$, then $\text{range}(Q_j) = \mathcal{K}_j(A, q_1)$ is an A-invariant subspace, the iteration stops, and we have solved the symmetric eigenvalue problem (21): $\Lambda(T_j) = \Lambda(A)$, $j = \text{rank}(A)$, and T_j is orthogonally similar to A. We summarize the motivating principles the Lanczos iteration with a Lemma.

Lemma 2 ([13, 14])**.** *Given a symmetric rank-r matrix $A \in \mathbb{R}^{n\times n}$ whose matrix-vector operator $A \mapsto Ax$ has complexity $O(\mathcal{M}(n))$ time, the Lanczos iteration computes $\Lambda(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ in $O(\mathcal{M}(n)r)$ time and $O(m)$ storage complexity, when computation is done in exact arithmetic.*

Even before an A-invariant subspace is obtained, intermediate spectral information is readily available in T_j, for $j < \text{rank}(A)$. Diagonalizing $T_j = Y\Theta Y^T$ yields value/vector pairs $\{(\theta_1^{(j)}, y_1^{(j)}), \ldots, (\theta_j^{(j)}, y_j^{(j)})\}$, called *Ritz pairs*. The values θ are called *Ritz values* and their associated vectors $v = Qy$ in the range of Q are called *Ritz vectors*. Ritz pairs satisfy $w^T(Ay - \theta y) = 0$ for all $w \in \mathcal{K}_j(A, q_1)$, and thus are intimately connected to the underlying Krylov subspace. From the Ritz perspective, the Lanczos iteration implicitly maintains two orthonormal basis for $K_j(A, q_1)$—a Lanczos basis Q and the Ritz basis Y:

$$A = QTQ^T = QY\Theta Y^T Q^T \iff AQY = QY\Theta$$

In principle, the Lanczos basis $\{q_i\}_{i=1}^j$ changes each iteration, while the Ritz basis $\{Qy_i^{(j)}\}_{i=1}^j$ changes after each subspace projection. The way in which the Ritz values approach the spectrum of A is well-studied [], as they are known to be Rayleigh-Ritz approximations of A's eigenpairs $\Lambda(A) = \{(\lambda_1, v_1), \ldots, (\lambda_j, v_j)\}$, and they are collectively known to be optimal in the sense that $T_k = B$ is the matrix that minimizes $\|AQ_k - Q_kB\|_2$ over the space of all $k \times k$ matrices. Moreover, Ritz values contain intrinsic information of the distance between $\Lambda(T_j)$ and $\Lambda(A)$. To see this, note that:

$$\|Av_i^{(j)} - v_i^{(j)}\theta_i^{(j)}\| = \beta_i^{(j)} = \beta_{j+1} \cdot |\langle e_j, y_i^{(j)}\rangle| \tag{27}$$

Thus, we need not necessarily keep the Lanczos vectors Q in memory to monitor how close the spectra of the T_j's approximate $\Lambda(A)$. In fact, it is known that the Ritz values $\{\theta_1^{(1)}, \theta_1^{(2)}, \ldots, \theta_1^{(j)}\}$ of T_j satisfy:

$$|\lambda - \theta_i^{(j)}| \leq \left(\beta_i^{(j)}\right)^2 / \left(\min_{\mu} |\mu - \theta_i^{(j)}|\right) \tag{28}$$

The Lanczos iteration's widespread adoption is in part due to its computational simplicity at reaching tridiagonal form: each step consists of a single matrix-vector product and a few vector operations. Moreover, the three-term recurrence from (26) implies the Lanczos iteration may be carried out with just three $O(n)$-sized vectors. Thus, if A is sparse and/or has special structure, the iteration may be carried in far fewer than $\approx n^3$ operations and with far less than the $\approx n^2$ memory—the typical time and storage complexities exhibited by direct methods. Indeed, for any symmetric $A \in \mathbb{R}^{n\times n}$ rank-r matrix with an average of ν nonzeros per row, approximately $(2\nu + 8)n$ flops are needed for a single Lanczos step, implying a $O(n\nu r)$ time complexity for a single iteration [10].

The Lanczos iteration and its many variants are part of a family of so-called "matrix free" methods—from the interfacing side, obtaining an eigen-decomposition of a symmetric real matrix A requires only a $v \mapsto Av$ operator. In particular, since A is not modified at all during the computation, one need necessarily involve explicitly storing A in memory. The sparsity pattern ∂_p inherits from the simplicial structure of K confers a few computational advantages compared to more general linear operators. Observe each column of ∂_p corresponds to a p-chain of the form (**??**), which is a constant up to difference in sign. Thus, evaluating $y = \partial_p^T v$ reduces to computing the locations $(k_1, k_2, \ldots, k_{p+1})$ (how to phrase this?):

$$y_i = y_i \pm S_{i,j}(\sigma_k, \sigma_l) \tag{29}$$

In particular, the operator $v \mapsto \partial_p \partial_p^T v$ need not be represented explicitly in memory: as long as we have access to the underlying simplicial complex K and its corresponding filter function f, we may take advantage of the . To make this more precise, we start with a useful lemma:

Lemma 3. *Let K denote a simplicial complex with (m, n) simplices of dimension $(p, p-1)$. Assume $m > n$. For any fixed $p \geq 0$ and arbitrary vector $v \in \mathbb{R}^n$, the operation $v \mapsto \partial_p \partial_p^T v$ can be evaluated $O(m)$ time and $O(m)$ memory.*

Proof. □

From this lemma, we have the following result:

Proposition 3 ([13, 14])**.** *Given a simplicial complex K with $m = |K^{(p+1)}|$ and $n = |K^{(p)}|$. Without loss of generality, assume $n < m$, and that* $\operatorname{rank}(\partial_p) = r$*. Then the Lanczos iteration from [] yields the quantity:*

$$\dim(B_p(K;\mathbb{R})) = \operatorname{rank}(\partial_{p+1}(K)) \tag{30}$$

in $O(mr)$ time and $O(m)$ storage complexity, when computation is done in exact arithmetic.

A proof of this proposition is given in the appendix.

For expository purposes, the preceding sections assumed exact arithmetic.

The structure of $L_p = \partial_p \partial_p^T$ is actually amenable to more than just computational efficiency in terms of matrix-vector multiplication. A matrix $A = \{a_{ij}\}$ is said to be *diagonally dominant* (DD) if it satisfies:

$$a_{ii} \geq \sum_{j \neq i} |a_{ij}|\,, \quad \forall\, i, j \in [n] \tag{31}$$

Similarly, when the inequality in (31) is strict, A is said to be *strictly diagonally dominant* (SDD). It is well known that for any $p \geq 0$, L_p is the smallest diagonally dominant matrix in the sense that $L_p[i,i] = \sum_{j \neq i} |L_p[i,j]|$. Diagonally dominant matrices are of particular interest becsue.... their spectrums can be approximated rather quickly.

An $O(m)$ `matvec` operator: Given a graph $G = (V, E)$ with edge set $E \subseteq V \times V$, let $A \in \mathbb{R}^{n \times n}$ denote its *adjacency matrix* with non-zero entries $A[i,j] = 1$ for every $(i,j) \in E$, and 0 otherwise. The *graph Laplacian* L is defined as:

$$L = D - A = \partial_1 \partial_1^T \tag{32}$$

where D is a diagonal matrix whose entries $d_{ii} = \{\, \deg(v_i) \,\}$ are the vertex degrees $\deg(v_i) = \sum_{j \neq i} A[i,j]$. The graph Laplacian is heavily structured matrix that is known to capture the connectivity structure of G. In particular, the structure of the graph Laplacian imply linear and quadratic forms which have particular graph interpretations:

$$(\,\forall\, x \in \mathbb{R}^n\,) \qquad (Lx)_i = \deg(v_i) \cdot x_i - \sum_{i \sim j} x_j\,, \qquad x^T L x = \sum_{i \sim j} (x_i - x_j)^2 \tag{33}$$

where we use the notation $i \sim j$ to indicate vertices $i, j \in V$ are path-connected in G. Moreover, L is symmetric, positive semi-definite, and the connection between L and ∂_1 via (32) suggests the elementary chain modifications from (16) can be adapted to the linear and quadratic forms from (33). Consider boundary matrices ∂_1^E and ∂_1^V whose chain entries are parameterized by edge and vertex functions, $f_E : V \times V \to \mathbb{R}$ and $f_v : V \to \mathbb{R}$, respectively:

$$\partial_1^E[k,l] = \begin{cases} \pm\, f_E(\sigma_l) & \text{if } \sigma_k \in \partial_p(\sigma_l) \\ 0 & \text{otherwise} \end{cases}, \qquad \partial_1^V[k,l] = \begin{cases} \pm\, f_V(\sigma_k) & \text{if } \sigma_k \in \partial_p(\sigma_l) \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

Both parameterizations generate two kinds of weighted graph Laplacians via (32). In the edge-weighted case, the Laplacian matrix-vector form is given by:

$$\left((\partial_1^E)(\partial_1^E)^T x\right)_i = (L^E x)_i = x_i \cdot d_f(v_i) - \sum_{i \sim j} x_j \cdot f_E(v_i, v_j)^2 \tag{35}$$

where $d_f(v_i) = \sum_{j \sim i} f_E(v_i, v_j)^2$. In the vertex-weighted case, it is given by:

$$\left((\partial_1^V)(\partial_1^V)^T x\right)_i = (L^V x)_i = x_i \cdot d(v_i) \cdot f_V(v_i)^2 - \sum_{i \sim j} x_j \cdot f_V(v_i) \cdot f_V(v_j) \tag{36}$$

Ignoring sign cancellations, the (i,j)-th entry of the Laplacian matrix L^{VE} formed by combining (32) with the hadamard product $\partial_1^{VE} = \partial_1^V \circ \partial_1^E$ has the following form:

$$L^{VE}[i,j] = \left((\partial_1^{VE})(\partial_1^{VE})^T\right)[i,j] = \begin{cases} d_f(v_i) \cdot f_V(v_i)^2 & i = j \\ -(f_E(v_i, v_j)^2 \cdot f_V(v_i) \cdot f_V(v_j)) & i \sim j \\ 0 & i \nsim j \end{cases} \tag{37}$$

In all cases, if $L = \partial_1 \partial_1^T$ where $\partial_1 \in \mathbb{R}^{n \times m}$ and $n < m$, then one may observe that the corresponding matrix-vector product for equations (35)-(37) can all be evaluated in $O(m)$ time.

4 Continuous Rank Relaxation

5 Approximation & Perturbation theory

The stability of the spectral relaxation proposed in (**??**) lends some stability the computation of β and $\hat{\beta}$ in the parameterized setting.

Namely, if (K, f_h) is a parameterized filtration where $f : \mathcal{H} \times K \to \mathbb{R}$ is a real-valued filter function that various smoothly in $\mathcal{H}$, one would expect the spectra of the constitutive terms in β_p^* and μ_p^* to also vary smoothly as a function of $\mathcal{H}$.

Let $\hat{\partial}_h$

6 Applications

Directional Transform: The canonical interpretation of the information displayed by a persistence diagram is that is summarizes the persistence of the sublevel sets of filtered space. Given a filtration pair (K, f) where K is a finite simplicial complex and $f : K \to \mathbb{R}$ is a real-valued function, the sublevel sets $|K|_i = f^{-1}(-\infty, i]$ deformation retract to... If K is embedded in $\mathbb{R}^d$, then geometrically f takes on the interpretation of a 'height' function whose range yields the 'height' of every simplex in K.

Let $X \subset \mathbb{R}^d$ denote a data set which can be written as a finite simplicial complex K whose simplices are PL-embedded in $\mathbb{R}^d$. Given this setting, define the *directional transform* (DT) of K as follows:

$$\begin{aligned} \mathrm{DT}(K) : S^{d-1} &\to K \times C(K, \mathbb{R}) \\ v &\mapsto (K_\bullet, f_v) \end{aligned}$$

where we write $(K_\bullet, f)$ to indicate the filtration on K induced by f_v for all $\alpha \in \mathbb{R}$, i.e.:

$$K_\bullet = K(v)_\alpha = \{\, x \in X \mid \langle x, v \rangle \leq \alpha \,\} \tag{38}$$

Conceptually, we think of DT as an S^{d-1}-parameterized family of filtrations.

The Persistent Homology Transform (PHT) is a shape statistic that establishes a fundamental connection between the topological information summarized by K's PH groups and the geometry of its associated embedding. Given a complex K built from X, it is defined as:

$$\begin{aligned} \mathrm{PHT}(K) : S^{d-1} &\to \mathcal{D}^d \\ v &\mapsto \big(\mathrm{dgm}_0(K, v), \mathrm{dgm}_1(K, v), \ldots, \mathrm{dgm}_{d-1}(K, v) \big) \end{aligned} \tag{39}$$

where $\mathcal{D}$ denotes the space of p-dimensional persistence diagrams, for all $p = 0, \ldots, d-1$ and S^{d-1} the unit $d-1$ sphere. The stability of persistence diagrams ensures that the map $v \mapsto \mathrm{dgm}_p(K, v)$ is Lipschitz with respect to the bottleneck distance metric $d_B(\cdot, \cdot)$ whenever K is a finite simplicial complex. Thus, the PHT may be thought of as an element in $C(S^{d-1}, \mathcal{D}^d)$: .

The primary result of [] is that the PHT is injective on the space of subsets of R^d that can be written as finite simplicial complexes[1], which we denote as $\mathcal{K}_d$. Equivalently, $\mathcal{K}_d$ decomposes space of all pairs (K, f) under the equivalence $(K, f) \sim (K, f')$ when $f(K) = f'(K)$.

A Appendix

B Laplacian Interpretation

We adapt an example from [12] to better convey the information carried by the Laplacian form of the boundary matrices.

Example B.1 (Adapted from [12])**.** Suppose the vertices of G are ordered and labeled from 1 to n arbitrarily such that, given any subset $X \subseteq V$, we may define column vector $x = (\, x_i \,)$ whose components $x_i = 1$ indicate $i \in X$ and $x_i = 0$ otherwise. Given such a set $X \subseteq V$, let $X' = V \setminus X$ denote its complement set. By L's definition, we have:

$$\begin{aligned} (Lx)_i > 0 &\iff i \in X \text{ and } |c_i(X)| = (Lx)_i \\ (Lx)_i < 0 &\iff i \in X' \text{ and } |c_i(X')| = |(Lx)_i| \\ (Lx)_i = 0 &\iff i \in X \cup X' \text{ and } c_i(X) = \emptyset \end{aligned}$$

where $c_v(X) = \{(v, w) \in E \mid v \in X \text{ and } w \in V \setminus X\}$ denotes the *cutset* of X restricted to v, i.e. the set of edges having as one endpoint $v \in X$ and another endpoint outside of X.

In other words, example B.1 demonstrates that L captures exactly how X is connected to the rest of G. Notice that if $X = V$, then $Lx = 0$ and thus 0 must be an eigenvalue of L with an eigenvector pair $\mathbf{1}$. Like the adjacency matrix, the interpretation of the matrix-vector product has a natural extension to powers of L, wherein just as entries in A^k model paths, entries in L^k are seen to model boundaries [12].

Proofs:

[1] Implicit in the injectivity statement of the PHT is that, given a subset $X \subset \mathbb{R}^d$ which may be written as finite simplicial complex K, the restriction $f : X \to \mathbb{R}$ to any simplex in K must is linear.

Proof of Lemma 1

Proof. The Pairing Uniqueness Lemma [8] asserts that if $R = \partial V$ is a decomposition of the total $m \times m$ boundary matrix ∂, then for any $1 \leq i < j \leq m$ we have $\text{low}_R[j] = i$ if and only if $r_\partial(i,j) = 1$. As a result, for $1 \leq i < j \leq m$, we have:

$$\text{low}_R[j] = i \iff r_R(i,j) \neq 0 \iff r_\partial(i,j) \neq 0 \tag{40}$$

Extending this result to equation (12) can be seen by observing that in the decomposition, $R = \partial V$, the matrix V is full-rank and obtained from the identity matrix I via a sequence of rank-preserving (elementary) left-to-right column additions. □

Proof of Proposition 1

Proof. We first need to show that $\beta_p^{i,j}$ can be expressed as a sum of rank functions. Note that by the rank-nullity theorem, so we may rewrite (10) as:

$$\beta_p^{i,j} = \dim\left(C_p(K_i)\right) - \dim\left(B_{p-1}(K_i)\right) - \dim\left(Z_p(K_i) \cap B_p(K_j)\right)$$

The dimensions of groups $C_p(K_i)$ and $B_p(K_i)$ are given directly by the ranks of diagonal and boundary matrices, yielding:

$$\beta_p^{i,j} = \text{rank}(I_p^{1,i}) - \text{rank}(\partial_p^{1,i}) - \dim\left(Z_p(K_i) \cap B_p(K_j)\right)$$

To express the intersection term, note that we need to find a way to express the number of p-cycles born at or before index i that became boundaries before index j. Observe that the non-zero columns of R_{p+1} with index at most j span $B_p(K_j)$, i.e $\{\, \text{col}_{R_{p+1}[k]} \neq 0 \mid\ k \in [j] \,\} \in \text{Im}(\partial_{p+1}^{1,j})$. Now, since the low entries of the non-zero columns of R_{p+1} are unique, we have:

$$\dim(Z_p(K_i) \cap B_p(K_i)) = |\Gamma_p^{i,j}| \tag{41}$$

where $\Gamma_p^{i,j} = \{\, \text{col}_{R_{p+1}[k]} \neq 0 \mid\ k \in [j],\, 1 \leq \text{low}_{R_{p+1}}[k] \leq i \,\}$. Consider the complementary matrix $\bar{\Gamma}_p^{i,j}$, given by the non-zero columns of R_{p+1} with index at most j that are not in $\Gamma_p^{i,j}$, i.e. the columns satisfying $\text{low}_{R_{p+1}}[k] > i$. Combining rank-nullity with the observation above, we have:

$$|\bar{\Gamma}_p^{i,j}| = \dim(B_p(K_j)) - |\Gamma_p^{i,j}| = \text{rank}(R_{p+1}^{i+1,j}) \tag{42}$$

Combining equations (41) and (42) yields:

$$\dim(Z_p(K_i) \cap B_p(K_j)) = |\Gamma_p^{i,j}| = \dim(B_p(K_j)) - |\bar{\Gamma}_p^{i,j}| = \text{rank}(R_{p+1}^{1,j}) - \text{rank}(R_{p+1}^{i+1,j}) \tag{43}$$

Observing the final matrices in (43) are *lower-left* submatrices of R_{p+1}, the final expression (13) follows by applying Lemma 1 repeatedly. □

Proof of boundary matrix properties

Proof. First, consider property (1). For any $t \in T$, applying the boundary operator ∂_p to $K_t = \text{Rips}_\epsilon(\delta_\mathcal{X}(t))$ with non-zero entries satisfying (**??**) by definition yields a matrix ∂_p satisfying $\text{rank}(\partial_p) = \dim(\text{B}_{p-1}(K_t))$. In contrast, definition (1) always produces p-boundary matrices of Δ_n; however, notice that the only entries which are non-zero are precisely those whose simplices σ that satisfy $\text{diam}(\sigma) < \epsilon$. Thus, $\text{rank}(\partial_p^t) = \dim(\text{B}_{p-1}(K_t))$ for all $t \in T$. $<$ (show proof of (2))$>$ Property (3) follows from the construction of ∂_p and from the inequality $\|A\|_2 \leq \sqrt{m}\|A\|_1$ for an $n \times m$ matrix A, as $\|\partial_p^t\|_1 \leq (p+1)\,\epsilon$ for all $t \in T$.

□

Dynamic Metric Spaces: Consider an $\mathbb{R}$-parameterized metric space $\delta_X = (X, d_X(\cdot))$ where X is a finite set and $d_X(\cdot) : \mathbb{R} \times X \times X \to \mathbb{R}_+$, satisfying:

1. For every $t \in \mathbb{R}, \delta_X(t) = (X, d_X(t))$ is a pseudo-metric space[2]
2. For fixed $x, x' \in X$, $d_X(\cdot)(x,x') : \mathbb{R} \to \mathbb{R}_+$ is continuous.

When the parameter $t \in \mathbb{R}$ is interpreted as *time*, the above yields a natural characterization of a "time-varying" metric space. More generally, we refer to an $\mathbb{R}^h$-parameterized metric space as *dynamic metric space*(DMS). Such space have been studied more in-depth [] and have been shown...

[2]This is required so that if one can distinguish the two distinct points $x, x' \in X$ incase $d_X(t)(x,x') = 0$ at some $t \in \mathbb{R}$.

Application: Time-varying : Let $\delta_{\mathcal{X}}$ denote an T-parameterized metric space $\delta_{\mathcal{X}}(\cdot) = (X, d_X(\cdot))$, where $d_X : \mathrm{T} \times X \times X \to \mathbb{R}_+$ is called a *time-varying metric* and X is a finite set with fixed cardinality $|X| = n$. δ_X as called a *dynamic metric space* (DMS) iff $d_X(\cdot)(x, x')$ is continuous for every pair $x, x' \in X$ and $\delta_{\mathcal{X}}(t) = (X, d_X(t))$ is a pseudo-metric space for every $t \in \mathrm{T}$. For a fixed $t \in \mathrm{T}$, the Rips complex at scale $\epsilon \in \mathbb{R}$ is the abstract simplicial complex given by

$$\mathrm{Rips}_\epsilon(\delta_{\mathcal{X}}(t)) := \{\sigma \subset X : d_X(t)(x, x') \leq \epsilon \text{ for all } x, x' \in \sigma\} \tag{44}$$

As before, the family of Rips complexes for varying $\epsilon > 0$ yields a filtration whose inclusion maps induce linear maps at the level of homology. The time-varying counterpart is analogous. In this context, we write the p-th persistent Betti number with respect to fixed values $i, j \in I$ as a function of $t \in \mathrm{T}$:

$$\beta_p^{i,j}(t) = \left(\dim \circ \mathrm{H}_p^{i,j} \circ \mathrm{Rips} \circ \delta_{\mathcal{X}}\right)(t) \tag{45}$$

References

[1] Ulrich Bauer. Ripser: efficient computation of vietoris–rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423, 2021.

[2] 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.

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

[4] Chao Chen and Michael Kerber. An output-sensitive algorithm for persistent homology. In *Proceedings of the twenty-seventh annual symposium on Computational geometry*, pages 207–216, 2011.

[5] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[6] 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.

[7] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 119–126, 2006.

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

[9] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[10] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

[11] Ming Gu and Stanley C Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM Journal on Matrix Analysis and Applications*, 16(1):172–191, 1995.

[12] Michael William Newman. The laplacian spectrum of graphs. Master's thesis, 2001.

[13] Beresford N Parlett. Do we fully understand the symmetric lanczos algorithm yet. *Brown et al*, 3:93–107, 1994.

[14] Horst D Simon. Analysis of the symmetric lanczos algorithm with reorthogonalization methods. *Linear algebra and its applications*, 61:101–131, 1984.

[15] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

A Boundary matrix factorization

Definition 2 (Boundary matrix decomposition)**.** *Given a filtration $K_\bullet$ with m simplices, let ∂ denote its $m \times m$ filtered boundary matrix. We call the factorization $R = \partial V$ the* boundary matrix decomposition *of ∂ if:*

I1. V is full-rank upper-triangular

I2. R satisfies $\mathrm{low}_R[i] \neq \mathrm{low}_R[j]$ iff its i-th and j-th columns are nonzero

where $\mathrm{low}_R(i)$ denotes the row index of lowest non-zero entry of column i in R or null *if it doesn't exist. Any matrix R satisfying property (I2) is said to be* reduced*; that is, no two columns share the same low-row indices.*

B Laplacian facts

In general, the spectrum of the graph Laplacian L is unbounded, [] and instead many prefer to work within the "normalized" setting where eigenvalues are bounded. The *normalized Laplacian* $\mathcal{L}$ of a graph G is typically given as:

$$\mathcal{L}(G) = D^{-1/2} L D^{-1/2} \tag{46}$$

with the convention that $D^{-1}(v_i, v_i) = 0$ for $\deg(v_i) = 0$. The variational characterization of eigenvalues in terms of the Rayleigh quotient of $\mathcal{L}$ convey a particular form. Specifically, for any real-valued function $f : V \to \mathbb{R}$ on G, when viewed as a column vector, $\mathcal{L}$ satisfies:

$$\frac{\langle f, \mathcal{L} f \rangle}{\langle f, f \rangle} = \frac{\sum\limits_{i \sim j} (g(v_i) - g(v_j))^2}{\sum\limits_{i} g(v_i)^2 \cdot \deg(v_i)} \tag{47}$$

where $f = D^{1/2} g$ and $\langle f, g \rangle$ denotes the standard inner product in $\mathbb{R}^n$. Equation (47) may be used to show that the spectrum $\Lambda(\mathcal{L})$ is bounded in the interval $[0, 2]$. In particular, it is known that:

$$\lambda_i \leq \sup_f \frac{\langle f, \mathcal{L} f \rangle}{\langle f, f \rangle} \leq 2 \tag{48}$$

Recall that, when G is connected, 0 is an eigenvalue of both L and $\mathcal{L}(G)$, with multiplicity $\mathrm{cc}(G)$. Moreover, if G is the union of disjoint graphs $G_1, G_2, \ldots, G_k$, then it has as its spectrum the union of the spectra $\Lambda(G_1), \Lambda(G_2), \ldots, \Lambda(G_k)$. Certain parts of the spectrum of $\mathcal{L}$ can be deduced explicitly for very structured types of G, such as complete graphs, complete bipartite graphs, star graphs, path graphs, and cycle graphs, and n-cubes. For a list of additional properties the graph and normalized Laplacians satisfy, including bounds on eigenvalues, relation to random walks and rapidly-mixing Markov chains, identities tied to isoperimetric properties of graphs, and explicit connections to spectral Riemannian geometry, see [5] and references within.