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 how this PBN characterization admits certain computational advantages in both static and parameterized settings. Namely we show the p-th PBN of a filtered simplicial complex $K_\bullet$ which 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 4.

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 as 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]. Thus, without loss of generality, we exclusively consider essential simplexwise filtrations and for brevity refer to them simply 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:

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

where $\hat{v}_i$ indicates the removal of v_i from the ith summand. The p-boundary of a p-chain is defined linearly in terms of its constitutive simplices. A p-chain with zero boundary is called a p-cycle, and together they form $Z_p(K) = \operatorname{Ker} \partial_p$. Similarly, the collection of p-boundaries forms $B_p(K) = \operatorname{Im} \partial_{p+1}$. 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 valid pairs of filtration indices. 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) [15] demarcating the evolution of homology classes: 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

Both the persistent homology (PH) groups and their corresponding persistent Betti numbers (PBNs) have long been studied from multiple, equivalent perspectives by several authors, see [2, 3, 6, 15] and references therein for an overview. 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. They also give a cubic-time algorithm to compute these groups on spaces in arbitrary dimensions over any field. 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) demonstrates an intrinsic connection between the multiplicity function and the persistent Betti numbers. Indeed, the fundamental lemma of persistent homology [9] states that for every pair of indices $0 \leq k \leq l \leq n+1$:

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

The customary interpretation of (7) is that persistence diagrams completely characterize their PBNs. This perspective has by and large driven the computational pipeline: if one has a space of interest $\mathcal{X}$ and is interested in computing its PBNs $\beta_p^{i,j}$ for any choice of $i < j$, the standard approach is to first compute $\mathrm{dgm}_p(\mathcal{X})$ such that the corresponding PBNs can be read-off directly. Nonetheless, the inverse interpretation via (7) has certain advantages. Its computation depends completely on the choice of indices (i, j): if $|i-j|$ is small, the computation implies a localized computation. Moreover, if one has computed $\beta_p^{i,j}$ for some choice of (i, j), the inner loop of the reduction algorithm from [9] suggests $\beta_p^{i+1,j}$ or $\beta_p^{i,j+1}$ can be determined quadratic time. Indeed, if given diagram has a finite number of off-diagonal points, then in theory for (7) to hold it must be the case that one can recover the diagram with a finite number of PBN computations. Indeed, this fundamental observation lead [4] to create a divide-and-conquer like PH algorithm, wherein positional information about the diagram is accessed solely through μ-queries—combinations of PBN-like computations.

After introducing several alternative notions of 'tameness', [3] generalize (6) using measure-theoretic tools to prove the existence and stability of persistence diagram. In particular, for M a decomposable persistence module over $\mathbb{R}$, Chazal [3] show that one may recover its persistence diagram by constructing its corresponding *persistence measure*:

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

The essential perspective from (8) is that the multiplicity function μ_p^* can be viewed as an integer-valued measure over rectangles in the plane. 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 (Theorem 3.11) expressing the persistent Betti number function $\beta_* : \Delta_+ \to \mathbb{N} \cup \{\infty\}$ as a sum of multiplicity functions, akin to (7). 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, justifying their use and study in continuously parameterized settings.

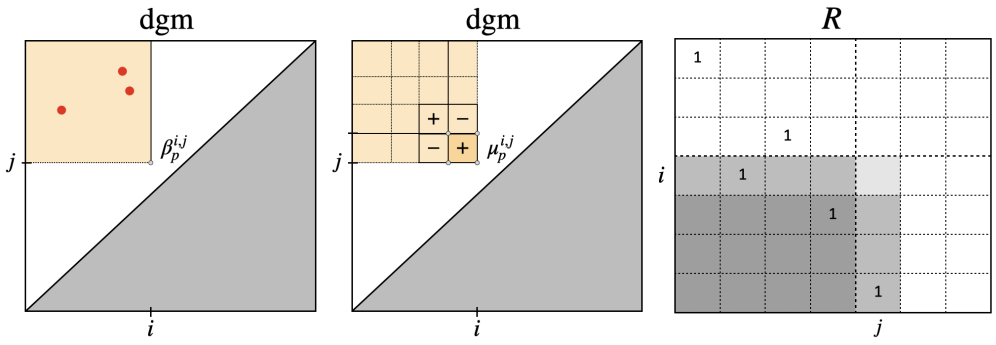

Figure 1: (Left) the persistent Betti number $\beta_p^{i,j}$ counts the number of points (3) in upper left-corner of $\mathrm{dgm}_p(K_\bullet)$. (Middle) The additivity of PBNs can be used to express multiplicity $\mu_p^{i,j}$ of any given box. (Right) The computational interpretation of the Pairing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

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 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 some index $(i,j) \in \Delta_+^m$ is defined as:

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

While $Z_p(K_i) = \mathrm{nullity}(\partial_p(K_i))$ and is thus easily obtained, efficient computation of the intersection term (the persistence part) is a bit more subtle. Zomorodian et al [15] give a procedure 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. However, the standard homology computations require $O(m^2)$ space and $O(m^3)$ time to compute, making the PBN computation no more efficient than 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 result 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. In fact, Lemma 1 may be further generalized to arbitrary rectangles in Δ_+ via *μ-queries* [4]: box-parameterized rank-based queries that count the number of persistence pairs that intersect a fixed "box" placed in the upper half-plane. Our first result of this effort we show is that this Lemma allows us to write the persistent Betti number as a sum of rank functions.

Proposition 1. *For any fixed $p \geq 0$, let ∂_p denote the p-dimensional boundary matrices of filtration $K_\bullet$ of size $n = |K^{(p)}|$, $m = |K^{(p+1)}|$. For any pair $(i,j) \in ([n],[m])$, the persistent Betti number $\beta_p^{i,j}$ at (i,j) is given by:*

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

where $I_p^{1,i}$ denotes the first i columns of the $n \times n$ identity matrix.

A detailed proof using Lemma 1 is given in the appendix. The main utility this proposition provides is that it enables the persistent Betti number as a combination of rank computations performed directly on the *unfactored* dimension p and $(p+1)$ boundary matrices. We dedicate the rest of the paper to exploring the consequences of this fact.

To get some intuition on what the structure and size of these matrices, we include a picture of each of the terms in Equation (13). Boundary matrices are sparse and highly structured: given $K_\bullet$ with m simplices $\sigma_1, \sigma_2, \ldots, \sigma_m$

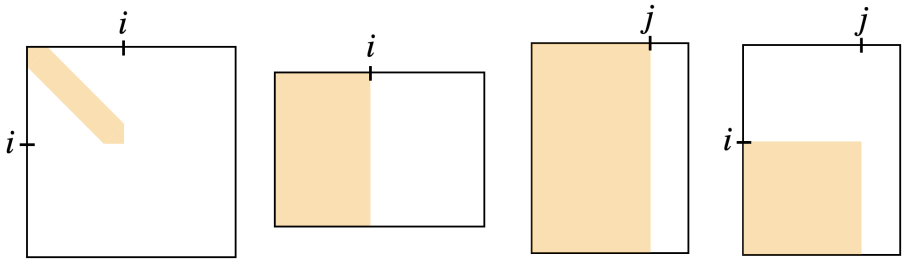

Figure 2: The four matrices whose ranks yield $\beta_p^{i,j}$ in the same order as given in (13). Each solid portion represents (sparse) blocks of non-zero entries, while each white portion is zero. Observe $\partial_{p+1}^{i+1,j} \subseteq \partial_{p+1}^{1,j}$ can be obtained by intersecting the non-zero entries of $\partial_{p+1}^{1,j}$ with the non-zero entries in the complement of $\partial_p^{1,i}$.

constructed from a p-dimensional complex K, its full boundary matrix ∂ is upper-triangular and has a storage complexity of:

$$\operatorname{nnz}(\partial) \sim O(m \log m) \tag{14}$$

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

Corollary 1. *Given a filtration $K_\bullet$ of size $n = |K_i^{(p)}|$ and $m = |K_j^{(p+1)}|$ and two indices $i \in [n]$, $j \in [m]$, computing $\beta_p^{i,j}$ can be done in time and storage complexity $O(\max\{\, R_\partial(n,i,p),\, R_\partial(m,j,p+1)\,\})$ where $R_\partial(a,b,c)$ is the complexity of computing the rank of a c-dimensional $a \times b$ boundary matrix with $b \cdot (c+1)$ 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: One advantage of expressing the PBN via equation (13) is that certain properties of rank function may be exploited in *parameterized* settings, i.e. settings where the scalar-valued filter function and corresponding filtration belong to a parametric family. One simple such property of the rank function is permutation invariance: given any square matrix $A \in \mathbb{R}^{n \times n}$, $\operatorname{rank}(A) = \operatorname{rank}(P^T A P)$ for any permutation matrix P. Observe the boundary matrices ∂_p in (13) need not be in filtration order induced by $(K_\bullet, f)$ to be evaluated—one may permute the matrices in (13) and obtain the same Betti number, so long as the matrices have the same essential non-zero pattern. Another approach to tracking the PBNs in parameterized setting is through the vineyards algorithm [7], however this requires $O(m^2)$ memory and an $\approx O(m^2)$ preprocessing procedure to detect changes in the filtration order[1] to simulate persistence across a homotopy, the canonical continuous parameterized

[1]The bound $O(m^2)$ assumes the homotopy changes each filtration value monotonically throughout the homotopy. Otherwise the number of order changes is clearly unbounded.

setting. In contrast, as we show below, PBNs need no such preprocessing procedure and may be computed in effectively $O(m)$ memory, even in parameterized settings. We elaborate in the following section.

Recall that the boundary operator ∂_p for a filtration pair $(K_\bullet, f)$ with $m = |C_p(K_\bullet)|$ and $n = |C_{p-1}(K_\bullet)|$ can be represented by an ordered $(n \times m)$ boundary matrix ∂_p whose columns and rows correspond to p-simplices and $(p-1)$-simplices. After orienting K in an arbitrary but consistent way, the entries of ∂_p have the form:

$$\partial_p[k,l] = \begin{cases} c(\sigma_j) & \text{if } \sigma_l \in \partial_p(\sigma_k) \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where $c(\sigma_*) \in \mathbb{F}$ is an arbitrary constant satisfying $c(\sigma) = -c(\sigma')$ if σ and σ' are opposite orientations of the same simplex, typically set to ± 1. In what follows, we assume a fixed orientation is given on K, and write $\pm c(\sigma)$ to indicate the sign of $c(\sigma)$ depends on the orientation of σ. Note the ordering of ∂_p must respect the facet poset of $K_\bullet$: if $\tau, \sigma \in K_\bullet^{(p)}$ and $f(\tau) \leq f(\sigma)$, then the chain $\partial_p(\tau)$ must appear before $\partial(\sigma)$.

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 you wanted to compute $\beta_p^{i,j}$ over $\mathcal{H}$. We give several application settings where this kind of formulation occurs naturally in section 4. To simplify this process, we introduce the notion of a *parameterized boundary matrix*.

Definition 1 (Parameterized boundary matrix)**.** *Let X denote a data set of interest of size $|X| = n$, equipped with parameterized filtering function $f : \mathcal{P}(X) \times \mathcal{H} \to \mathbb{R}$. Define $(\mathcal{P}(X), \preceq^*)$ be a fixed linear extension of the face poset of the standard $(n-1)$-simplex. For fixed $i, j \in \mathbb{F}$, define the* $\mathcal{H}$-parameterized p-th boundary matrix $\hat{\partial}_p^{i,j}(h)$ *at scale (i,j) to be the $\binom{n}{p} \times \binom{n}{p+1}$ matrix ordered by $\preceq^*$ for all $h \in \mathcal{H}$, and whose entries (k,l) satisfy:*

$$\hat{\partial}_p^{i,j}(h)[k,l] = \begin{cases} \pm\, S_{i,j}(\sigma_k, \sigma_l) & \textit{if } \sigma_k \in \partial_p(\sigma_l) \\ 0 & \textit{otherwise} \end{cases} \tag{16}$$

where $S_{i,j} : \mathcal{P}(X) \times \mathcal{P}(X) \to \{0,1\}$ is a step *function that accepts a face/coface pair (σ_k, σ_l) and returns a 1 if $f(\sigma_k) \geq i$ and $f(\sigma_l) \leq j$, and 0 otherwise.*

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 $r(A) = \operatorname{rank}(A)$.

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

One remark is in order: note that although definition 1 specifies $\hat{\partial}_p^{i,j}$ as a full $\binom{n}{p} \times \binom{n}{q}$ matrix, implying a memory complexity of $O(q \cdot n^q)$ for all $h \in \mathcal{H}$, we remark that there is no need to fully allocate this much memory as the rows/columns corresponding to the set of p/q face/coface pairs (σ_k, σ_l) with $f(\sigma_l) < i$ or $f(\sigma_k) > j$ are entirely 0. As we will show below, it is enough to have access to the simplices $K_j^{(q)}(h)$.

Laplacian Connection: Due to its connection as the discrete analogue of the Laplace differential operator, the graph Laplacian and its spectrum have been studied in many areas of mathematical research as they often have particular geometric interpretations in physical chemistry, and financial mathematics []. Since the investigation of the question "can you hear the shape of a drum?" by [], the question of what geometric properties of a manifold are determined by its Laplace operator has been of pertinent interest to practitioners and theoreticians alike.

Given a graph $G = (V, E)$ with edge set $E \subseteq V \times V$, let $A \in \mathbb{R}^{n \times n}$ denote its corresponding *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:

$$L = D - A = \partial_1 \partial_1^T \tag{18}$$

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 and its many variants [] are well studied. Structurally, they admit a number are attractive features: L is symmetric and thus its eigenvalues are real-valued; L is positive semi-definite ($L \succeq 0$) and thus satisfies $x^T L x \geq 0$ for all $x \neq 0$; $\operatorname{rank}(L)$ is given by $n - \operatorname{cc}(G)$, where $\operatorname{cc}(G)$ is the number of connected components of G (and thus, $0 \in \Lambda(L)$ with multiplicity $\operatorname{cc}(G)$). In particular, the linear and quadratic forms L admits have particular graph interpretations:

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

where we use the notation $i \sim j$ to indicate vertices $i, j \in V$ are path-connected in G. The graph Laplacian is generally known to capture the connectivity structure of G. To make this more concrete, we adapt an example 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 and $c_v(X) = \{(v, w) \in E \mid v \in X \text{ and } w \in V \setminus X\}$ its *cutset* restricted to v, i.e. the set of edges having as one endpoint $v \in X$ and another endpoint outside of X. By L's definition, we have:

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

In other words, L captures exactly how X is connected to the rest of G. Notice that if $X = V$, then $y = 0$ and thus 0 must be an eigenvalue of L. The interpretation extends can be generalized to powers of L, wherein just as entries in A^k model paths, entries in L^k are seen to model boundaries [12].

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 and carry a particular form. Define the *normalized Laplacian* $\mathcal{L}$ of a graph G as:

$$\mathcal{L}(G) = D^{-1/2} L D^{-1/2} \tag{19}$$

with the convention that $D^{-1}(v_i, v_i) = 0$ for $\deg(v_i) = 0$. Note that $\mathcal{L}(G)$ can also be written in a form $\partial_1 \partial_1^T$ by replacing entries column entries $(1, -1)$ corresponding to edges $e = (u, v)$ with $(1/\sqrt{\deg(\mathrm{u})}, -1/\sqrt{\deg(\mathrm{v})})$ [**?**]. 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:

$$\frac{\langle f, \mathcal{L} f \rangle}{\langle f, f \rangle} = \frac{\sum_{i \sim j} (g(v_i) - g(v_j))^2}{\sum_i g(v_i)^2 \cdot \deg(v_i)} \tag{20}$$

where $f = D^{1/2} g$ and $\langle f, g \rangle$ denotes the standard inner product in $\mathbb{R}^n$. Equation (20) may be used to show that the spectrum $\Lambda(\mathcal{L})$ is bounded in the interval $[0, 2]$. In particular, it is known that:

$$\lambda_i \leq \sup_f \frac{\langle f, \mathcal{L} f \rangle}{\langle f, f \rangle} \leq 2 \tag{21}$$

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.

The graph laplacian appears naturally in the computation of β_0^*. Consider the rank expression from (13). Since ∂_0 is trivial, the expression reduces to:

$$\beta_p^{i,j} = |K_i^{(0)}| - \mathrm{rank}(\partial_1^j) + \mathrm{rank}(\partial_1^{i+1,j}) \tag{22}$$

Since $\mathrm{rank}(A) = \mathrm{rank}(AA^T) = \mathrm{rank}(A^T A)$, we may equivalently express the second and third terms in terms of their graph Laplacians. Let $G_j = \mathcal{L}(K_j)$ denote the graph of K_j...

For $p \geq 1$, the PBN expression can still be generalized via *up Laplacians*. The higher-dimensional generalizations of the graph Laplacian for simplicial complexes have been studied in a variety of settings, see e.g. []. In particular, given K a finite oriented simplicial complex and some $p \geq 0$, the p-th *combinatorial Laplacian* of K is the linear operator $\mathcal{L}_p : C_p \to C_p$ given by:

$$\mathcal{L}_p = \partial_{p+1} \circ \partial_{p+1}^T + \partial_p^T \circ \partial_p \tag{23}$$

For convenience, we use the notation $\uparrow L_p = \partial_{p+1} \circ \partial_{p+1}^T$ and $\downarrow L_p = \partial_p^T \circ \partial_p$ to denote the so-called *up* and *down* combinatorial p-th Laplacians, respectively. With this notation, notice the computation of the PBN can be expressed via the ranks of up Laplacians...

Complexity of Persistence:

Rank Computation: In this section we discuss a few 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, where applicable, how these methods made be adapted to efficiently compute (17).

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 are the spaces spanned by:

$$\mathcal{K}_j(A, v) := \text{span}\{v, Av, A^2v, \ldots, A^{j-1}\} = \text{range}(K_j(A, v)) \tag{24}$$

where $K_j(A, v) = [v \mid Av \mid A^2v \mid \cdots \mid A^{j-1}]$ are the corresponding Krylov matrices. The motivation for Krylov subspaces comes 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 $x = A^{-1}v \in K_m(A, v)$ and $K_m(A, v)$ is an invariant subspace of A. This observation combined with the fact A is orthogonally diagonalizable (since A is symmetric) implies 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 $K_j(A, v) = Q_jR_j$ for each $j = 1, 2, \ldots, n$. Due to A's symmetry and the orthogonality of Q_j, we have $q_k^T Aq_l = q_l^T A^T q_k = 0$ for $k > l + 1$, implying the corresponding $T_j = Q_j^T AQ_j$ have a tridiagonal structure:

$$T_j = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{j-1} \\ & & & \beta_{j-1} & \alpha_j \end{bmatrix}, \ \beta_j > 0, \ j = 1, 2, \ldots, n \tag{25}$$

Unlike the spectral decomposition $A = U\Lambda U^T$, which identifies a diagonalizable A with its spectrum $\Lambda(A)$ up to a change of basis $A \mapsto M^{-1}AM$, there is no canonical choice of T_j due to the arbitrary choice of v. Instead, the Lanczos method exploits a particular 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:

$$K_n(A, q_1) = QQ^T K_n(A, q_1) = Q[\, e_1 \mid Te_1 \mid T^2e_1 \mid \cdots \mid T^{n-1}e_1 \,] \tag{26}$$

is the QR factorization of $K_n(A, q_1)$. 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:

$$AQ_j = Q_jT_j + \beta_j q_{j+1} e_j^T \tag{27}$$

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 (27) yields the three-term recurrence:

$$Aq_j = \beta_{j\text{-}1}\, q_{j\text{-}1} + \alpha_j\, q_j + \beta_j\, q_{j+1} \tag{28}$$

where $\alpha_j = q_j^T Aq_j$, $\beta_j = \|r_j\|_2$, $r_j = (A - \alpha_j I)q_j - \beta_{j\text{-}1}q_j$, and $q_{j+1} = r_j/\beta_j$. Observe that if $(q_{j\text{-}1}, \beta_j, q_j)$ are known, then $(\alpha_j,\ \beta_{j+1}, q_{j+1})$ are completely determined and (28) may be used iteratively to obtain T_j—this sequential process is called the *Lanczos iteration*. One may begin the iteration by setting β_0 and q_0 to 0, and q_1 arbitrarily. 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 still an A-invariant subspace, the iteration stops, and we have $j = \text{rank}(A)$ and $\Lambda(T_j) = \Lambda(A)$.

The Lanczos iteration's widespread adoption is in part due to its simplicity: each step consists of just a matrix-vector operator and a few vector operations. It is often referred to as a "black box" method due to the fact that a fast $v \mapsto Av$ operator is the only aspect of the computation that explicitly depends on A. If A is sparse and/or has special structure, then the Lanczos method can be quite efficient. Indeed, the three-term recurrence from (28) implies the Lanczos iteration of A may be carried out with just three $O(n)$-sized vectors. 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]. In general, eigenvalue computations consist of two phases, of which the reduction to structured form (the first phase) is the bottleneck: once T_j has been obtained for sufficient $j > 0$, the corresponding spectrum $\Lambda(T_j)$ may be obtained in $O(n \log n)$ time [11].

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 (15), which is a constant up to difference in sign. Thus, evaluating $y = \partial_p^T v$ reduces to computing the locations $(k_1, k_2, \ldots, k_{p+1})$:

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

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 2. *For any fixed $p \geq 0$, given simplicial complex K with m (n, respectively) simplices of dimension p ($p-1$, respectively) and an arbitrary vector v of size $|v| = n$, the operation $v \mapsto \partial_p \partial_p^T v$ can be evaluated $O(\max\{m, n\})$ time and $O(\max\{m, n\})$ memory.*

From this lemma, we have the following result:

Proposition 2 ([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.

Convergence Rate: In principle, if the Lanczos iteration doesn't break down due to rounding errors, then the characteristic polynomial of T_k is the unique polynomial $p^k \in P^k$ that achieves []:

$$\|p^k(A)b\| = minimum \tag{31}$$

In fact, there is more we can say: if $S = \mathcal{K}_k(A, q_1)$, then applying the Lanczos iteration yield the *Ritz pairs* (θ, y) of A, which are the pairs satisfying $w^T(Ay - \theta y) = 0$ for all $w \in S$. The essential result from [] is that the eigenvalues $\Lambda(T_k) = \{\theta_1, \theta_2, \ldots, \theta_k\}$ are 'optimal' in the sense that $T_k = B$ is the matrix that minimizes $\|AQ_k - QkB\|_2$ over the space of all $k \times k$ matrices. Thus, if $\operatorname{rank}(A) = r$, the eigenvalues of T_k for some $k < r$ tend to converge quickly to the extremal eigenvalues of A. The iteration typically converges geometrically and can be stopped as soon as the desired accuracy is reached.

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:

$$a_{ii} \geq \sum_{j \neq i} |a_{ij}|, \quad \forall\, i, j \in [n] \tag{32}$$

Similarly, when the inequality in (32) 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.

4 Applications

Consider an $n \times m$ boundary matrix ∂_p representing the boundary operator $C_p(K) \to C_{p-1}(K)$. Our goal is to have the ability to compute the ranks of submatrices of ∂_p as efficiently as possible, possibly in a parameterized stream.

In general, the complexity of computing the rank of a $n \times m$ matrix A depends heavily on both the structure of A (e.g. it's sparsity) as well as the choice of coefficients $\mathbb{F}$. Typical persistence computations are carried out using coefficients in $\mathbb{Z}/2\mathbb{Z}$ as the computation is greatly simplified in this setting. Indeed, the persistence computation is not invariant to the choice of field [].

Here, we consider the complexity of computing $\beta_p^{i,j}$ with real-valued coefficients. Since real numbers may not be representable explicitly in finite precision, we will consider the *numerical rank*.

Definition 2 (Numerical Rank)**.** *Let $A \in \mathbb{R}^{m \times n}$ have singular values $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_r(A) > 0$. For some $0 < \epsilon < 1$, the ϵ-rank of A, denoted by $\text{rank}_\epsilon(A)$, is the smallest integer k such that:*

$$\text{rank}_\epsilon(A) = \min_{k \geq 0} \{\, k : \sigma_{k+1}(X) \leq \epsilon \cdot \|A\|_2 \,\} \tag{33}$$

The most direct way to compute (33) is compute the singular values of a matrix up to working precision, however this requires $O(mn^2)$ complexity in general [].

It is instructive to consider Observe that computation of β_p^* involves rank computations on subsets of the $p, p+1$ boundary matrices. When $p = 0$, we have:

$$\begin{aligned}
\beta_0^{i,j} &= \text{rank}(I_p^{1,i}) - \text{rank}(\partial_p^{1,i}) - \text{rank}(\partial_{p+1}^{1,j}) + \text{rank}(\partial_{p+1}^{i+1,j}) \\
&= |C_0(K_i)| - \text{rank}_\epsilon((\partial_1^{1,j})(\partial_1^{1,j})^T) + \text{rank}_\epsilon((\partial_1^{i+1,j})(\partial_1^{i+1,j})^T) \\
&= |C_0(K_i)| - \text{rank}_\epsilon(L_1^j) + \text{rank}_\epsilon((\partial_1^{i+1,j})(\partial_1^{i+1,j})^T)
\end{aligned}$$

Where L_1^j is a graph Laplacian the subgraph of K consisting of only edges $e \in K_1$ with $f(e) \leq j$.

A Appendix

Proofs:

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:

$$\text{low}_R[j] = i \iff r_R(i,j) \neq 0 \iff r_\partial(i,j) \neq 0 \tag{34}$$

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:

$$\dim(Z_p(K_i) \cap B_p(K_i)) = |\Gamma_p^{i,j}| \tag{35}$$

where $\Gamma_p^{i,j} = \{\, \mathrm{col}_{R_{p+1}[k]} \neq 0 \mid k \in [j],\, 1 \leq \mathrm{low}_{R_{p+1}}[k] \leq i \,\}$. 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 $\mathrm{low}_{R_{p+1}}[k] > i$. Combining rank-nullity with the observation above, we have:

$$|\bar{\Gamma}_p^{i,j}| = \dim(B_p(K_j)) - |\Gamma_p^{i,j}| = \mathrm{rank}(R_{p+1}^{i+1,j}) \tag{36}$$

Combining equations (35) and (36) yields:

$$\dim(Z_p(K_i) \cap B_p(K_j)) = |\Gamma_p^{i,j}| = \dim(B_p(K_j)) - |\bar{\Gamma}_p^{i,j}| = \mathrm{rank}(R_{p+1}^{1,j}) - \mathrm{rank}(R_{p+1}^{i+1,j}) \tag{37}$$

Observing the final matrices in (37) 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 $K_t = \mathrm{Rips}_\epsilon(\delta_{\mathcal{X}}(t))$ with non-zero entries satisfying (15) by definition yields a matrix ∂_p satisfying $\mathrm{rank}(\partial_p) = \dim(\mathrm{B}_{p-1}(K_t))$. 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 $\mathrm{diam}(\sigma) < \epsilon$. Thus, $\mathrm{rank}(\partial_p^t) = \dim(\mathrm{B}_{p-1}(K_t))$ 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...

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

$$\mathrm{Rips}_\epsilon(\delta_{\mathcal{X}}(t)) := \{\sigma \subset X : d_X(t)(x, x') \leq \epsilon \text{ for all } x, x' \in \sigma\} \tag{38}$$

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 $t \in \mathrm{T}$:

$$\beta_p^{i,j}(t) = \left(\dim \circ \mathrm{H}_p^{i,j} \circ \mathrm{Rips} \circ \delta_{\mathcal{X}}\right)(t) \tag{39}$$

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.

[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 $t \in \mathbb{R}$.

[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 3 (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.*