Math 583B: Topological Data Analysis

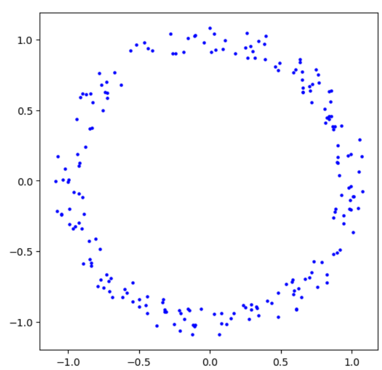
Kyle Ormsby

Spring 2024

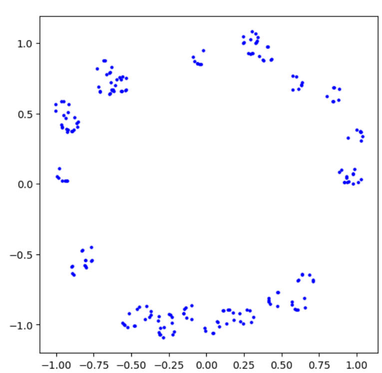
1 *Inferring the shape of data* — 25 March 2024

Imagine that you're running an experiment in which you measure a large number — say N — of real-valued variables with each observation. Each observation is then a point in $\mathbb{R}^N$, and if you make k total observations, then the data associated with your experiment is a *point cloud* $P = \{x_1, x_2, \ldots, x_k\} \subseteq \mathbb{R}^N$.

If the system being observed is not purely random, then — up to issues of noise and accuracy — we expect P to be sampled from a subspace $M \subseteq \mathbb{R}^N$. How might we infer the structure and shape of M from P, at least under the assumption that k is relatively large? This is one of the questions that topological data analysis (TDA) aims to answer, at least for particular notions of "structure" and "shape". In the figure presented here, we see a point cloud P in $\mathbb{R}^2$ sampled with noise from the unit circle $S^1 \subseteq \mathbb{R}^2$, and we seek algorithmic methods that will recognize (features of) S^1 as the underlying space from which P is sampled. Of course, in practice, N might be very large, and it is unlikely that your visual cortex will rise to the challenge of guessing M.



But even for small N, we can still ask more from our methods. Consider the displayed point cloud $Q \subseteq \mathbb{R}^2$ which exhibits strikingly different structure at different scales. At small scales, points seem to be sampled from disjoint circles. After zooming out (so at a larger scale), those small circles seem to assemble into one big copy of S^1. The tools we will develop are *scale independent* and do not depend on parameter tuning. We will ultimately produce concise, interpretable summaries that capture the nature of data at all scales.



Our first and primary tool will be the *persistent homology* of the *Čech* or *Vietoris–Rips filtered complex* associated with a point cloud $P \subseteq \mathbb{R}^N$. We may view this as a two-step process:

$$\begin{tikzcd} \{\text{point clouds}\} \arrow[r, bend left, "\check{\mathbb{C}}"] \arrow[r, bend right, "\mathbb{VR}"'] & \{\text{filtered spaces}\} \arrow[r, "\mathrm{PH}"] & \{\text{persistence modules}\}. \end{tikzcd}$$

A *filtered space* $\mathscr{X} = \{X_s\}_{s\in\mathbb{R}}$ is a collection of spaces[1] X_s indexed by scales $s \in \mathbb{R}$ such that

$$s \le t \implies X_s \subseteq X_t.$$

For the purposes of this introduction, we will focus on the Čech filtered complex $\check{\mathbb{C}}(P)$ of our point cloud $P \subseteq \mathbb{R}^N$. At scale $s \in \mathbb{R}$, $\check{\mathbb{C}}_s(P)$ is

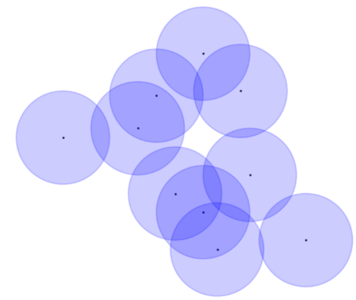
[1] By *space* we might mean topological space or (abstract) simplicial complex. If working with complexes, we take $\subseteq$ to mean subcomplex.

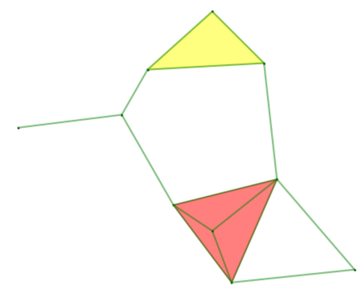
the simplicial complex with one n-simplex for each subset $A \subseteq P$ with $|A| = n+1$ and

$$\bigcap_{x \in A} \overline{B}_s(x) \neq \varnothing.$$

In other words, we get an n-simplex for each $(n+1)$-subset of P for which the closed Euclidean balls of radius s centered at points of A have nonempty common intersection. Since the intersection condition becomes less stringent as s gets larger, we have that $\check{\mathrm{C}}_s(P)$ is a subcomplex of $\check{\mathrm{C}}_t(P)$ when $s \leq t$. Later, we will encounter the Nerve Lemma, which roughly says that $\check{\mathrm{C}}_s(P)$ is homotopy equivalent to $\bigcup_{x \in P} \overline{B}_s(P)$ in reasonable scenarios. Note that the combinatorial nature of $\check{\mathrm{C}}_s(P)$ makes it much better adapted to computation than the filtered topological space $\{\bigcup_{x \in P} \overline{B}_s(P)\}_{s \in \mathbb{R}}$.

We write $\overline{B}_s(x)$ for the closed ball of radius s centered at x.





Black points are 0-simplices, green edges are 1-simplices, yellow shading is a 2-simplex, and red shading is a 3-simplex. Note that the bottom right triangle is not filled in yellow because the triple intersection of the balls around those vertices is empty.

Now that we have a filtered space $\mathscr{X} = \check{\mathrm{C}}(P)$, we aim to capture features of each space $X_s := \check{\mathrm{C}}_s(P)$ and how these features are related as the filtration parameter changes. Taking a cue from algebraic topology, we view $\mathrm{H}_*(X_s; \mathbb{F})$ — the homology[2] of X_s with coefficients in a field $\mathbb{F}$ — as a good summary of the features of X_s. Functoriality of homology then provides us with $\mathbb{F}$-linear transformations

$$(\iota_s^t)_* \colon \mathrm{H}_*(X_s; \mathbb{F}) \longrightarrow \mathrm{H}_*(X_t; \mathbb{F})$$

for $s \leq t$ and $\iota_s^t \colon X_s \subseteq X_t$, and these maps $(i_s^t)_*$ provide our comparisons of features. Packaging all of the homologies and comparisons maps together produces a *persistence module* $\mathrm{PH}_*(\mathscr{X}; \mathbb{F})$, the $\mathbb{F}$-*persistent homology* of $\mathscr{X}$, which is our scale independent summary of the shape of our data.

[2] We will review homology theory next lecture. It is a lie in the direction of truth to say that the dimension of the $\mathbb{F}$-vector space $\mathrm{H}_n(X_s; \mathbb{F})$ measures the number of n-dimensional "holes" in X_s.

The miracle here is that persistence modules admit a convenient and complete invariant called a *barcode* or (after a mild but tremendously beneficial transformation) *persistence diagram*. To give the flavor of barcodes, we will consider a simplified scenario in which we have $\mathbb{F}$-vector spaces $\{V_i\}_{i \in \mathbb{N}}$ and linear transformations $\iota_i^j \colon V_i \to V_j$ for $0 \leq i \leq j$ such that

(1) $\iota_i^i = \mathrm{id}_{V_i}$ for all i, and

(2) for $0 \leq i \leq j \leq k$, $\iota_j^k \circ \iota_i^j = \iota_i^k$.

The essential data here is of the form

$$V_0 \to V_1 \to \cdots \to V_i \to V_{i+1} \to \cdots$$

and we may view the persistence module $(\{V_i\}_{i \in \mathbb{N}}, \{\iota_i^j\}_{i \leq j})$ as a functor $\mathscr{V} = \{V_i\}_{i \in \mathbb{N}}$ from the category associated with the partially ordered set $(\mathbb{N}, \leq)$ to the category of $\mathbb{F}$-vector spaces and linear transformations. Such a persistence module might arise from a point cloud by considering Čech complexes at scales $s_0 < s_1 < \cdots$.

Let $\mathbb{F}[t]$ denote the ring of polynomials in variable t over $\mathbb{F}$, graded so that $|t| = 1$, and set

$$\Theta(\mathscr{V}) := \bigoplus_{i \in \mathbb{N}} V_i.$$

Then we may endow $\Theta(\mathscr{V})$ with the structure of a graded $\mathbb{F}[t]$-module by setting the action of the polynomial generator t to be

$$t \cdot (v_i)_{i\in\mathbb{N}} := (\iota_{i-1}^{i} v_{i-1})_{i\in\mathbb{N}}$$

where $v_{-1} := 0$. In fact, Θ is an equivalence of categories between $\mathbb{N}$-persistence modules and graded $\mathbb{F}[t]$-modules.[3]

[3] The inverse functor takes M_* to $\{M_i\}_{i\in\mathbb{N}}$ with ι_i^j given by multiplication by t^{j-i}.

A common capstone theorem of a first course in algebra is the classification of finitely generated modules over a principal ideal domain. A graded version of this theorem holds *mutatis mutandis*, and so it behooves us to understand which persistence modules correspond to finitely generated graded $\mathbb{F}[t]$-modules. Call a persistence module $\mathscr{V} = \{V_i\}_{i\in\mathbb{N}}$ *tame* when every V_i is finite-dimensional and ι_i^{i+1} is an isomorphism for sufficiently large i. One may prove that $\mathscr{V}$ is tame if and only if $\Theta(\mathscr{V})$ is finitely generated over $\mathbb{F}[t]$.

By the classification theorem for finitely generated graded modules over a PID, if $\mathscr{V}$ is tame then there are (essentially unique) integers $i_1, \dots, i_m, j_1, \dots, j_n, \ell_1, \dots, \ell_n$ and an isomorphism

$$\Theta(\mathscr{V}) \cong \bigoplus_{s=1}^{m} \Sigma^{i_s}\mathbb{F}[t] \oplus \bigoplus_{t=1}^{n} \Sigma^{j_t}\mathbb{F}[t]/(t^{\ell_t})$$

where Σ^r denotes a grading shift upwards by r.[4] Translating this into the world of persistence modules, we learn that every tame persistence module decomposes (essentially uniquely) as

[4] That is, $(\Sigma^r M_*)_s = M_{s-r}$

$$\mathscr{V} \cong \bigoplus_{j=0}^{N} \mathbb{I}[b_j, d_j]$$

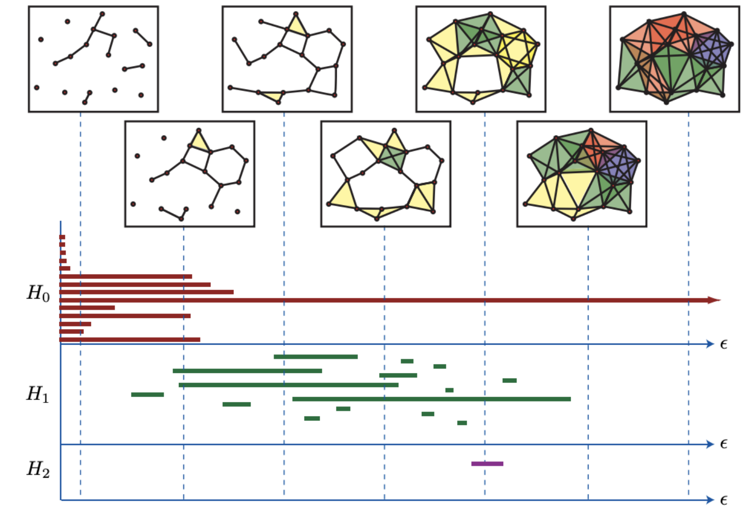
where each b_j is a nonnegative integer, $d_j \in \mathbb{N} \cup \{\infty\}$, and $\mathbb{I}[b_j, d_j]$ is the *interval persistence module* with

$$\mathbb{I}[b_j, d_j]_i = \begin{cases} \mathbb{F} & \text{if } b_j \leq i \leq d_j, \\ 0 & \text{otherwise,} \end{cases}$$

and $\iota_i^{i'} = \mathrm{id}_{\mathbb{F}}$ for $b_j \leq i \leq i' \leq d_j$.

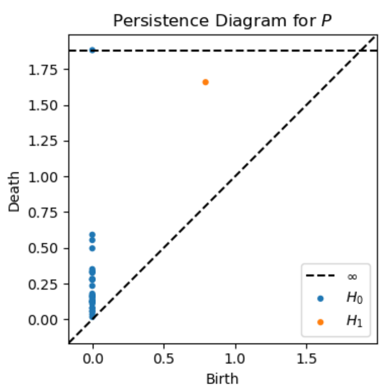
For an interval persistence module $\mathbb{I}[b, d]$, we refer to b as the *birth* and d as the *death* scale. We may then visualize the decomposition of $\mathscr{V}$ as a multiset of intervals $[b_j, d_j]$ called the *barcode* of $\mathscr{V}$. The following illustration is taken from Ghrist.[5] Beware, though, that it uses the Vietoris–Rips filtration instead of the Čech filtration; we will study $\mathbb{VR}$ in detail later.

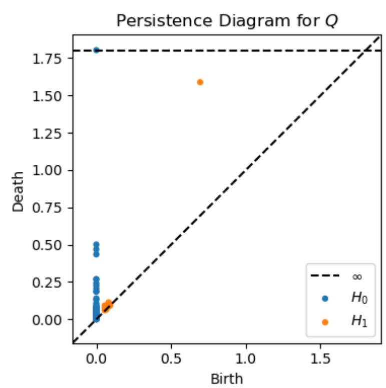
[5] Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75



While barcodes prevailed in the early days of TDA, experience has shown that *persistence diagrams* are better suited to statistical analysis. The persistence diagram of $\mathscr{V}$ consists of the multiset of points (b_j, d_j) lying on or above the diagonal of $\mathbb{N} \times (\mathbb{N} \cup \{\infty\})$.

The Vietoris–Rips filtered complexes of the data sets P and Q from our initial discussion have the following persistence diagrams (with PH_0 in blue and PH_1 in orange):





Focusing on the blue PH_0 classes, we see that in both cases all connected components are born at time 0, and at scales above ≈ 0.7 there is a single connected component that persists to $+\infty$. This last class is analogous to the red bar of infinite length in the previous diagram.

Looking at orange PH_1 classes, we can readily observe significant differences between the point clouds. In each, there is a highly persistent class born around scale 0.75, but Q detects the small scale structure as well, giving a cluster of short-lived PH_1 classes born around scale 0.1. These classes witness the small radii circles (arranged around the unit circle) from which Q is sampled.

It is often claimed that classes with large persistence $d - b$ (*i.e.*, those high above the diagonal) represent the "true" topology of the data, while small persistence classes correspond to noise. The point clouds P and Q illustrate that this is not necessarily the case.

1.1 Future topics

One of our primary tasks will be the development of pseudometrics allowing us to compare persistence diagrams. We leave this to future development, along with the many foundational details elided or overlooked in this introduction. Once the foundations are established, the rest of the course will focus on the following:

(1) applications of persistent homology to particular data modalities,

(2) extending persistent homology to filtrations indexed by more exotic partially ordered sets, and

(3) refining PH_0 via hierarchical clustering.

See the syllabus for a detailed (but flexible) schedule of topics.

1.2 Notes

The content of this introduction was primarily drawn from the Oudot's textbook[6] and Carlsson's survey article.[7] The original images were produced in Python using the Ripser persistent homology package.[8] We will use Ripser extensively when exploring examples and applications, and you should follow the installation instructions at `https://ripser.scikit-tda.org/` to get it working on your personal computer. You can find the Jupyter notebook used to produce diagrams from this and future lectures at `https://github.com/kyleormsby/math583`.

[6] Oudot, S. Y. (2015). *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI

[7] Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308

[8] Tralie, C., Saul, N., and Bar-On, R. (2018). Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925

1.3 Exercises

(1) Install the necessary software and run the demos from today's class on your personal computer.

(2) Determine the smallest[9] point cloud in $\mathbb{R}^3$ whose Čech filtered complex exhibits nonzero PH_2 as some scale. What about in $\mathbb{R}^2$?

[9] Smallest in terms of cadinality — the least number of points.

2 *Spaces, complexes, and homology* — 27 March 2024

2.1 Topology

From the Kleinian perspective, geometry is the study of properties invariant under isometries, that is, distance-preserving transformations. Indeed, when a geometer says that two triangles are the same (or *congruent* or *isometric*), they do not mean that each triangle consists of exactly the same points, but rather that one may translate, rotate, and reflect one triangle until it matches the other.

Felix Klein (1849–1925)

Topology plays a similar game, but with a much coarser notion of "sameness". We say that two spaces — the objects of topology — are *homeomorphic* when there are continuous functions between them that are mutually inverse. In this sense, topology is the study of properties that are invariant under homeomorphism. Such properties include such notions as connectivity and compactness, but exclude more rigid properties such as angle, distance, or volume.

We will generally assume that the reader is familiar with point-set topology, but will quickly recall some of the basic definitions.

The standard reference for point-set topology is Munkres; see also the recent graduate text of Bradley–Bryson–Terilla.

Munkres, J. R. (2000). *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, second edition; and Bradley, T.-D., Bryson, T., and Terilla, J. (2020). *Topology—a categorical approach*. MIT Press, Cambridge, MA

Definition 2.1. A *topological space* is a pair $(X, \mathscr{U})$ consisting of a set X and a collection of subsets $\mathscr{U} \subseteq 2^X$ called *open sets* such that

(1) $\varnothing$ and X are in $\mathscr{U}$,

(2) $\mathscr{U}$ is closed under arbitrary unions: $U_\alpha \in \mathscr{U}$ for $\alpha \in A$ implies $\bigcup_{\alpha\in A} U_\alpha \in \mathscr{U}$, and

(3) $\mathscr{U}$ is closed under finite intersections: $U_i \in \mathscr{U}$ for i in a finite set I implies $\bigcap_{i\in I} U_i \in \mathscr{U}$.

We will write X for $(X, \mathscr{U})$ when the topology $\mathscr{U}$ is clear from context. A subset $U \subseteq X$ is called *open* when it belongs to $\mathscr{U}$, and a subset $C \subseteq X$ is called *closed* when $X \smallsetminus C$ is open. These properties are not mutually exclusive, as exhibited by the *clopen* sets $\varnothing$ and X.

Example 2.2. In the *standard topology* on Euclidean space $\mathbb{R}^n$, a subset $U \subseteq \mathbb{R}^n$ is open if and only if it is a union of *open balls* $B_r(x) := \{y \in \mathbb{R}^n \mid |y - x| < r$. This is equivalent to saying that U is open if and only if for each $x \in U$ there exists $r > 0$ such that $B_r(x) \subseteq U$.

Example 2.3. Suppose X is a topological space and Y is a subset of X. We may endow Y with the *subspace* topology (relative to X) by declaring that the open sets of Y are exactly those sets of the form $U \cap Y$ for $U \subseteq X$ open.

As a subexample of subspaces, consider the interval $[0, 1] \subseteq \mathbb{R}$, where $\mathbb{R}$ carries the standard topology. Then $(1/2, 1] = (1/2, 3/2) \cap [0, 1]$ is open in $[0, 1]$, but not in $\mathbb{R}$.

Definition 2.4. A function $f\colon X \to Y$ between topologyical spaces is *continuous* when the preimage $f^{-1}U$ over every open set $U \subseteq Y$ is open in X. A continuous function $f\colon X \to Y$ is a *homeomorphism* when it admits a continuous inverse $g\colon Y \to X$. In this case, we say that X and Y are *homeomorphic* and write $X \cong Y$.

The categorically inclined reader will note that topological spaces and continuous functions form a category, and the isomorphisms in this category are exactly the homeomorphisms.

Example 2.5. If X is a topological space and $f\colon X \to \mathbb{R}$ is continuous, then the *sublevel set* $f^{-1}(-\infty, u) = \{x \in X \mid f(x) < u\}$ is open in X since the interval $(-\infty, u) = \{t \in \mathbb{R} \mid t < u\}$ is open in $\mathbb{R}$. Similarly, $f^{-1}(-\infty, u]$ is closed.[10]

[10] *Exercise*: For $f\colon X \to Y$ any continus map and $C \subseteq Y$ closed, check that $f^{-1}C$ is closed in X.

It will frequently be important to study a yet weaker notion of "sameness" in topology call *homotopy*. This is a two-step definition that first identifies when two continuous functions are homotopic, and then proceeds to spaces.

Definition 2.6. Continuous functions $f, g\colon X \to Y$ are *homotopic* when there exists a continuous function $H\colon X \times [0,1] \to Y$ such that the restriction of H to $X \times 0$ agrees with f and the restion to $X \times 1$ agrees with g. Such a map H is called a *homotopy* between f and g and we write $H\colon f \simeq g$.

We may think of H is a "movie" continuously interpolating between f and g.

Recall that spaces X and Y are homeomorphic when there are continuous functions $f\colon X \to Y$ and $g\colon Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. The following definition formalizes the notion a map admitting an inverse "up to homotopy".

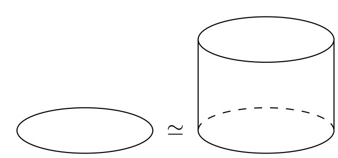
Definition 2.7. Two spaces X and Y are *homotopic* when there exist continuous functions $f\colon X \to Y$ and $g\colon Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$.

Example 2.8. We check that the circle S^1 is homotopic to the cylinder $S^1 \times [0,1]$. Take

$$\begin{aligned} f\colon S^1 &\longrightarrow S^1 \times [0,1] \\ z &\longmapsto (z, 0) \end{aligned}$$

and

$$\begin{aligned} g\colon S^1 \times [0,1] &\longrightarrow S^1 \\ (z,t) &\longmapsto z. \end{aligned}$$



Then $g \circ f = \mathrm{id}_{S^1}$ (which is clearly homotopic to id_{S^1}) and $f \circ g\colon (z,t) \mapsto (z, 0)$. We define

$$\begin{aligned} H\colon (S^1 \times [0,1]) \times [0,1] &\longrightarrow S^1 \times [0,1] \\ ((z,t), s) &\longmapsto (z, ts). \end{aligned}$$

We have $((z,t),0) \mapsto (z,0)$, which is $f \circ g$, while $((z,t),1) \mapsto (z,t)$, which is $\mathrm{id}_{S^1 \times [0,1]}$. Thus $H\colon f \circ g \simeq \mathrm{id}_{S^1 \times [0,1]}$, as needed to verify that $S^1 \simeq S^1 \times [0,1]$.

In fact, there was nothing special about S^1 in this argument. Any space X is homotopic to $X \times [0,1]$, the cylinder on X.

2.2 Geometric and abstract simplicial complexes

Let $P = \{x_0, \ldots, x_k\} \subseteq \mathbb{R}^N$ be a point cloud in $\mathbb{R}^N$. An *affine combination* of P is a sum of the form
$$\sum_{i=0}^{k} \lambda_i x_k$$
where $\lambda_i \in \mathbb{R}$ and $\sum \lambda_i = 1$. The collection of all affine combinations of P is called the *affine hull* of P; it is always an affine linear subspace of $\mathbb{R}^N$.

The point cloud P is *affinely independent* if no $x \in P$ is an affine combination of $P \smallsetminus \{P\}$. This is equivalent to the set $\{x_1 - x_0, \ldots, x_k - x_0\}$ being a linearly independent set of vectors.

Recall that a subset M of $\mathbb{R}^N$ is *convex* when the line segment joining any two points of M is a subset of M; the *convex hull* of M is the intersection $\operatorname{Conv}(M)$ of all convex sets containing M. When $P = \{x_0, \ldots, x_k\} \subseteq \mathbb{R}^N$ is affinely independent, we get *barycentric coordinates* on $\operatorname{Conv}(P)$. The barycentric coordinates of a point $x \in \operatorname{Conv}(P)$ are the unique $\lambda_i \in [0, 1]$ such that
$$x = \sum_{i=0}^{k} \lambda_i x_i \qquad \text{and} \qquad \sum_{i=0}^{k} \lambda_i = 1.$$

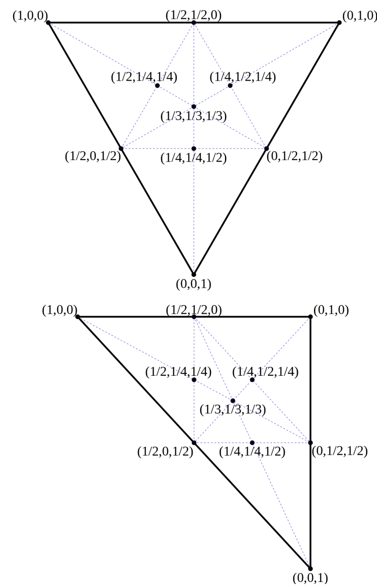


Image by user Rubybrian, CC BY-SA 3.0.

We can now define geometric simplicial complexes, whose basic building blocks are geometric simplices:

Definition 2.9. Suppose $k, N \in \mathbb{N}$ with $k \leq N$. A *geometric k-simplex* σ is the convex hull of an affinely independent point cloud $P = \{x_0, \ldots, x_k\} \subseteq \mathbb{R}^N$ with $k + 1$ elements, *i.e.*,
$$\sigma = \operatorname{Conv}(P).$$

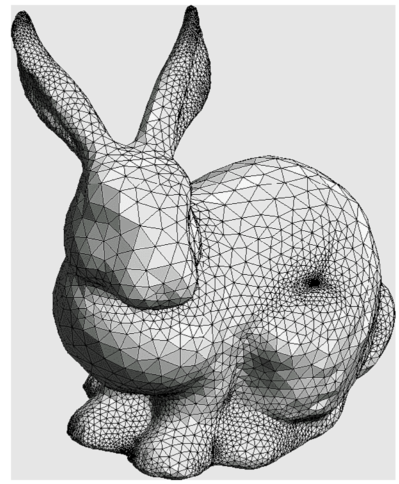
The *dimension* of σ is k, and we will sometimes write $\sigma = \sigma^k$ to express its dimension. The points $x_0, \ldots, x_k$ are the *vertices* of σ, its *edges* are the convex hulls of pairs of vertices of σ, and, more generally, the convex hull τ of any subset of P is called a *face* of σ. A face τ of σ is a *facet* when $\dim \tau = \dim \sigma - 1$.

Definition 2.10. Let $N \in \mathbb{N}$. A (finite) *geometric simplicial complex* $K \subseteq \mathbb{R}^N$ is a (finite) collection of geometric simplices K that is closed under taking faces ($\sigma \in K$ and τ a face of K implies $\tau \in K$) and compatible with intersection (if $\sigma, \tau \in K$, then $\sigma \cap \tau$ is either empty or a common face of both σ and τ).

The *dimension* of a geometric simplicial complex is the maximal dimension of its simplices. The *body* of K is
$$|K| = \bigcup_{\sigma \in K} \sigma.$$
In a standard act of laziness, we will often blur the distinction between K and $|K|$.

Definition 2.11. A *triangulation* of a subspace $X \subseteq \mathbb{R}^N$ is a geometric simplicial complex K in $\mathbb{R}^N$ such that $|K| \cong X$.

We warn the reader that not every subspace of $\mathbb{R}^N$ admits a triangulation. But "reasonable" spaces do, and there is a tremendous amount of interesting topology one can do with geometric simplicial complexes. Triangulations are also essential in computer graphics, as illustrated by the Stanford bunny pictured here.



While geometric simplicial complexes nicely match our intuition for how spaces might be chopped up into simplicial pieces, they are very inefficient as data structures. By working with abstract simplicial complexes, we can recover the homeomorphism type of (the body of) a geometric simplicial complex far more efficiently.

Definition 2.12. Let P be a finite set. An *abstract simplicial complex* L on P is a family of nonempty subsets of P that is closed under taking nonempty subsets: $\sigma \in L$ and $\varnothing \neq \tau \subseteq \sigma$ implies $\tau \in L$.

Example 2.13. Every geometric simplicial complex K on vertex set P determines an abstract simplicial complex L on P by declaring that $\sigma \in L$ if and only if $\operatorname{Conv}(\sigma)$ is a face of K.

We may also create a geometric simplicial complex from any abstract simplicial complex, a process called *geometric realization*. The defining feature of a geometric realization K of an abstract simplicial complex L is that the abstract simplicial complex L' associated with K (as in the above example) is L again up to relabeling of vertices. When P is finite, constructing geometric realizaitons is fairly straightforward:

Proposition 2.14. *Every abstract simplicial complex K with k vertices admits a geometric realization in $\mathbb{R}^{k-1}$.*

Proof. The complex K is a subcomplex of the full $(n-1)$-simplex on P, which may be geometrically realized as the convex hull of $0, e_1, \ldots, e_{n-1} \in \mathbb{R}^{n-1}$. □

It is actually the case that every abstract simplicial complex on k vertices admits a geometric realization in $\mathbb{R}^{2k+1}$, but we will not go into the proof. One can also geometrically realize infinite abstract simplicial complexes via a colimit construction.

In order to compare simplicial complexes, we need an appropriate class of maps.

Definition 2.15. A *simplicial map* between abstract simplicial complexes K and L is a function $f\colon K^{(0} \to L^{(0)}$ on vertices such that the image of every abstract simplex in K is an abstract simplex in L.

Definition 2.16. A *simplicial map* between geometric simplicial complexes K and L is a function $f\colon |K| \to |L|$ such that the restriction of f to $K^{(0)}$

induces a simplicial map between associated abstract simplicial complexes and which is linear on geometric simplices (in terms of barycentric coordinates), *i.e.*, if $t_0, \ldots, t_k \in [0,1]$ with $\sum t_i = 1$ and $v_0, \ldots, v_k \in K^{(0)}$, then

$$f\left(\sum t_i v_i\right) = \sum t_i f(v_i). \tag{2.17}$$

Tracing through the definitions, one may check that simplicial maps between geometric simplicial complexes induce abstract simplicial maps, and conversely, each simplicial map between abstract simplicial complexes extends in a unique way to a simplicial map between geometric realizations via (2.17). It is also the case that simplicial maps induce continuous functions between bodies. Perhaps surprising, though, is that continuous maps between (bodies of) geometric simplicial complexes can be approximated by simplicial maps.

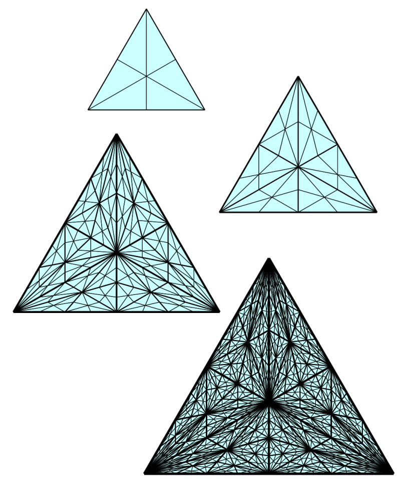


Image by user Studentofrationality showing successive barycentric subdivisions of an equilateral triangle, CC BY-SA 4.0.

Theorem 2.18 (Simplicial approximation)**.** *Suppose $f\colon K \to L$ is a continuous function between geometric simplicial complexes. Then there exist sufficiently fine subdivisions K' of K and L' of L, and a simplicial map $f'\colon K' \to L'$ such that $f \simeq f'$.*

Here a subdivision of K is a geometric simplicial complex K' such that every face of K is a union of simplices of K'. The proof of the simplicial approximation theorem is covered in many standard combinatorial or algebraic topology texts and we won't attempt it here.

We need one more crucial definition before we move on to simplicial homology, namely that of *orientation*. This amounts to a choice of ordering on vertices, taken up to a certain equivalence relation.

Definition 2.19. An *oriented simplex* on vertices $x_0, \ldots, x_k$ is an ordered $(k+1)$-tuple $\sigma = \langle x_0, x_1, \ldots, x_k \rangle$ subject to the rule

$$\sigma = \operatorname{sgn}(\pi)\langle x_{\pi(0)}, x_{\pi(1)}, \ldots, x_{\pi(k)} \rangle$$

where σ is a permutation of $\{0, 1, \ldots, k\}$ and $\operatorname{sgn}(\pi)$ is the *signature* of π:

$$\operatorname{sgn}(\pi) = (-1)^m$$

where m is the number of transpositions in a decomposition of σ as a composite of transpositions.[11]

[11] Such decompositions always exist; the number m is not unique, but all such m have the same parity.

We also give each 0-dimensional simplex two orientations, $\langle x \rangle$ and $-\langle x \rangle$.

Finally, two k-simplices sharing a $(k-1)$-dimensional face σ are *consistently oriented* when they induce opposite orientations on σ.

2.3 Simplicial homology

We are now going to "measure" the "holes" in a simplicial complex with a tool called homology. The slogan for homology is "cycles mod boundaries".

Both cycles and boundaries are special types of "chains", which are formal linear combinations of oriented simplices. The cycles are those chains with trivial boundary (so they properly "enclose" part of the complex), and the boundaries are boundaries of chains. Crucially, the boundary of a boundary is trivial, so every boundary chain is a cycle. The word "mod" means we take a quotient, an algebraic operation which identifies cycles when they differ by a common boundary. In particular, every boundary becomes 0. The idea here is that when the region enclosed by a cycle can be filled in (*i.e.*, the cycle is a boundary), it is *not* a hole. Meanwhile, cycles that don't bound a lower dimensional chain are essential and do get counted as holes.

We now formalize these ideas. Let $\mathbb{F}$ be a field,[12] let K be an abstract simplicial complex of dimension n, and let k be a natural number between 0 and n.

[12] That is, a number system in which you can add, multiply, and divide (by nonzero numbers). Examples include the rational, real, and complex numbers. We will extensively use the finite fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ which encode "clock arithmetc" on a clock with hours $0, 1, \ldots, p-1$ for p prime. Our favorite case will be $\mathbb{F}_2 = \{0, 1\}$ in which $2 = 0$ and $1 = -1$. Make sure you know what the addition and scalar multiplication rules for $C_k(K; \mathbb{F})$ are.

Definition 2.20. A *k-chain* is a formal $\mathbb{F}$-linear combination $\sum \lambda_i \sigma_i^k$ of oriented k-dimensional simplices in K subject to the rule $(-1) \cdot \sigma = -\sigma$, where the latter term indicates σ with the opposite orientation. The *chain group* $C_k(K; \mathbb{F})$ is the $\mathbb{F}$-vector space of all k-chains.

Note that if K has n_k many k-simplices, then $C_k(K; \mathbb{F}) \cong \mathbb{F}^{n_k}$ with basis given by the k-simplices.

We next formalize the notion of boundary. The rough idea is that the boundary of an k-simplex consists of all its $(k-1)$-dimensional facets added together (as a $(k-1)$-chain) with consistent orientations. We then extend this assignment linearly to all of $C_k(K; \mathbb{F})$. Given an oriented simplex $\sigma = \langle x_0, x_1, \ldots, x_k \rangle$ and $0 \le i \le k$, let

$$\hat{\sigma}_i := \langle x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k \rangle$$

be the $(k-1)$-dimensional facet of σ with x_i removed with the induced orientation.

Definition 2.21. Let $k \in \mathbb{N}$. The *boundary map*

$$\partial = \partial_k \colon C_k(K; \mathbb{F}) \longrightarrow C_{k-1}(K; \mathbb{F})$$

is the unique $\mathbb{F}$-linear transformation such that

$$\partial_k \sigma = \sum_{i=0}^{k} (-1)^i \hat{\sigma}_i.$$

If $k = 0$, we set $\partial_0 \colon C_0(K; \mathbb{F}) \to 0$ to be the trivial map.

If our combinatorial algebra correctly captures geometric intuition, then the boundary of a boundary should be trivial. This brings us to what Dennis Sullivan[13] calls the most important equation in mathematics: $\partial^2 = 0$.

[13] Sullivan won the 2022 Abel Prize for his contributions to algebraic topology, geometric topology, and dynamics.

Theorem 2.22. *For* $k \ge 1$,

$$\partial^2 := \partial_{k-1} \circ \partial_k = 0.$$

Proof. It suffices to show that $\partial^2\sigma = 0$ for $\sigma = \langle x_0, \ldots, x_k\rangle$ an oriented k-simplex. For $0 \le i < j \le k$, let $\hat{\sigma}_{ij}$ be the $(k-2)$-dimensional simplex with x_i and x_j removed. This term appears twice in the expansion of $\partial^2\sigma$, once with sign $(-1)^i(-1)^j$ and once with sign $(-1)^i(-1)^{j-1}$. These terms cancel so the total sum for $\partial^2\sigma$ is equal to 0. □

It follows that the image of ∂_k is contained in the kernel of ∂_{k-1}. This make the chain groups and ∂ into a *chain complex*:

$$\cdots \xrightarrow{\partial} C_n(K;\mathbb{F}) \xrightarrow{\partial} C_{n-1}(K;\mathbb{F}) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1(K;\mathbb{F}) \xrightarrow{\partial} C_0(K;\mathbb{F}) \xrightarrow{\partial} 0.$$

[TODO: Add example with matrices representing boundary maps; observe that consecutive matrices multiply to 0.]

We are now ready to implement our slogan of "homology equals cycles mod boundaries".

Definition 2.23. Let K be an abstract simplicial complex, let $\mathbb{F}$ be a field, and let k be a natural number. Then

» the group of *k-cycles*, $Z_k(K;\mathbb{F})$, is the kernel of ∂_k, which is a subspace of $C_k(K;\mathbb{F})$,

» the group fo *k-boundaries*, $B_k(K;\mathbb{F})$, is the image of ∂_{k+1}, which is a subspace of $Z_k(K;\mathbb{F})$, and

» the *k-th homology* group of K is the quotient vector space[14]

$$\mathrm{H}_k(K;\mathbb{F}) := Z_k(K;\mathbb{F})/B_k(K;\mathbb{F}).$$

We call $\dim H_k(K;\mathbb{F})$ the q-th $\mathbb{F}$-Betti number of K, denoted $b_q(K;\mathbb{F})$.

One amazing feature of homology is that it is both a homemorphism and homotopy invariant. Note that this is wildly false for chains, cycles, and boundaries. Implicit in this claim is that homology is also *functorial*: given a simplicial map $f\colon K \to L$, the assignment

$$\begin{aligned} f_*\colon H_k(K;\mathbb{F}) &\longrightarrow H_k(L;\mathbb{F}) \\ [\sigma] &\longmapsto \begin{cases} [f(\sigma)] & \text{if } f(x_0), \ldots, f(x_k) \text{ are distinct,} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

is a well-defined linear transformation. Moreover, $(\mathrm{id}_K)_* = \mathrm{id}_{H_k(K;\mathbb{F})}$ and if $g\colon L \to M$ is another simplicial map, then $(g \circ f)_* = g_* \circ f_*$. Homotopy invariance is now the following statement:

Theorem 2.24. *If $k \in \mathbb{N}$ and $f\colon K \simeq L$ is a simplicial homotopy equivalence, then $f_*\colon H_k(K;\mathbb{F}) \cong H_k(L;\mathbb{F})$.*

In particular, if K is *contractible*,[15] then $H_0(K;\mathbb{F}) \cong \mathbb{F}$ and $H_k(K;\mathbb{F}) = 0$

[14] If U is a vector supspace of V, then V/U is the vector space of U-*cosets* of the form $v + U = \{v + u \mid u \in U\}$ for $v \in V$. Note that if $v - w \in U$, then $v + U = w + U$. Addition is given by $(v+U) + (w+U) = (v+w) + U$ and scalar multiplication by $\lambda(v+U) = (\lambda v) + U$. If you're not familiar with quotient vector spaces, you should check that these operations are well-defined. Observe that $U = 0 + U$ is the trivial (or zero) element of V/U. This is the sense in which V/U "kills" the subspace U.
The quotient space V/U also enjoys a universal property. First note that there is a canonical quotient map $q\colon V \to V/U$ taking v to $v + U$. If $f\colon V \to W$ is a linear transformation such that $f(U) = 0$ – *i.e.*, $U \le \ker f$ – then there is a unique linear transformation $\tilde{f}\colon V/U \to W$ such that $f = \tilde{f} \circ q$. We say that linear transformation killing U *factor* uniquely through V/U (via q).

[15] *I.e.*, K is homotopy equivalent to a point, written $K \simeq *$.

for $k > 0$ simply by a quick computation of chains, cycles, and boundaries for a point.

We now list some of the many properties one would prove about simplicial homology in a full development of the subject:

» $b_0(K; \mathbb{F})$ is the number of connected components of K.

» Let S^n denote the unit sphere in $\mathbb{R}^{n+1}$. For $n \geq 1$,

$$H_k(S^n; \mathbb{F}) \cong \begin{cases} \mathbb{F} & \text{if } k = 0, n, \\ 0 & \text{otherwise.} \end{cases}$$

» Let $T^n := S^1 \times \cdots \times S^1$ be the n-fold Cartesian product of the circle with itself. Then

$$b_k(T^n; \mathbb{F}) = \binom{n}{k}$$

for all fields $\mathbb{F}$.

» Let $K \amalg L$ denote the disjoint union of simplicial complexes K and L. Then

$$H_k(K \amalg L; \mathbb{F}) \cong H_k(K; \mathbb{F}) \oplus H_k(L; \mathbb{F}).$$

Crucially, homology also has excellent properties with respect to decompositions into subcomplexes. The so-called *Mayer–Vietoris sequence* is a powerful method for computing the homology of $A \cup B$ in terms of the homology of A, B, and $A \cap B$ when A and B are subcomplexes of $A \cup B$. We can think of this tool as a "derived" version of the inclusion-exclusion theorem. Developing and even stating the theorem requires a fair bit of homological algebra, so we will point the reader to Section 8.2 of Virk's notes[16] for details.

[16] Virk, Ž. (2022). Introduction to persistent homology. https://zalozba.fri.uni-lj.si/virk2022.pdf. Accessed on 19 March 2024

2.4 Notes

The presentation of simplicial complexes and homology is a compressed version of Chapters 3 and 7 along with Section 4.2 of Virk's notes. I strongly recommend this text for those new to the subject!

2.5 Exercises

(1) Use chains, cycles, and boundaries to compute the homology of a circle, modeled as the simplicial set $\{a, b, c, ab, bc, ca\}$. (Here we are writing x for $\{x\}$ and xy for $\{x, y\}$.)

(2) Triangulate the Klein bottle K and prove that

$$H_k(K; \mathbb{F}_2) \cong \begin{cases} \mathbb{F}_2 & \text{if } k = 0, 2, \\ \mathbb{F}_2^2 & \text{if } k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

while

$$H_k(K;\mathbb{F}) \cong \begin{cases} \mathbb{F} & \text{if } k = 0,1, \\ 0 & \text{otherise} \end{cases}$$

if $\mathbb{F}$ is a field in which $2 \neq 0$. This demonstrates that homology is sensitive to the arithmetic of the field of coefficients!

(3) Define the *Euler characteristic* of a finte simplicial complex K with n_k many k-simplices to be the alternating sum

$$\chi(K) = \sum_{k\geq 0}(-1)^k n_k.$$

Use a rank-nullity argument to prove that $\chi(K)$ is also equal to the alternating sum of Betti numbers $\sum_{k\geq 0} b_k(K;\mathbb{F})$. Conclude (a) that Euler characteristic is a homotopy invariant and (b) that the alternating sum of Bett numbers does not depend on $\mathbb{F}$. (Observe that this is consistent with the $\mathbb{F}$-Betti numbers of the Klein bottle, which has Euler characteristic 0.)

3 *Persistence modules — 1 April*

Note: We didn't cover homology during our last meeting, so the 1 April lecture started with oriented simplices.

Recall fromt he first lecture the fundamental pipeline for persistent homology of point clouds:

$$\begin{tikzcd} \{\text{point clouds}\} \arrow[r, bend left, "\check{\mathrm{C}}"] \arrow[r, bend right, "\mathbb{VR}"'] & \{\text{filtered spaces}\} \arrow[r, "\mathrm{PH}"] & \{\text{persistence modules}\}. \end{tikzcd}$$

Given a filtered abstract simplicial complex $\mathscr{X} = \{X_s\}_{s\in\mathbb{R}}$ (with $\iota_s^t\colon X_s \subseteq X_t$ for $s \leq t$), we may now apply the k-th homology functor $H_k(-;\mathbb{F})$ to get a persistence module

$$\mathrm{PH}_k(\mathscr{X};\mathbb{F}) := \{H_k(X_s;\mathbb{F}), (\iota_s^t)_* \mid s \leq t \in \mathbb{R}\}.$$

Our current goal is to understand the structure of such persistence modules.

Definition 3.1. An $\mathbb{R}$-persistence module $\mathscr{V}$ over a field $\mathbb{F}$ is a collection of $\mathbb{F}$-vector spaces V_s, $s \in \mathbb{R}$, along with linear transition maps $\iota_s^t\colon V_s \to V_t$ for each pairs $s \leq t$ such that

(1) for $s \in \mathbb{R}$, $\iota_s^s = \mathrm{id}_{V_s}$, and

(2) for $r \leq s \leq t$, $\iota_s^t \circ \iota_r^s = \iota_r^t$.

A *map* of persistence modules $f\colon \mathscr{V} \to \mathscr{W}$ is a collection of linear transformations $f_s\colon V_s \to W_s$ such that the diagram

$$\begin{tikzcd} V_s \arrow[r, "f_s"] \arrow[d, "\iota_s^t"'] & W_s \arrow[d, "\iota_s^t"] \\ V_t \arrow[r, "f_t"'] & W_t \end{tikzcd}$$

commutes for all $s \leq t$. (Here we are abusing notations and writing ι_s^t for the transition maps for both $\mathscr{V}$ and $\mathscr{W}$.) Two persistence modules are *isomorphic*, written $\mathscr{V} \cong \mathscr{W}$, when there is a map of persistence modules $f\colon \mathscr{V} \to \mathscr{W}$ which admits a two-sided inverse or, equivalently, has f_s a bijection for each $s \in \mathbb{R}$.

Here's a categorical take on persistence modules: For any poset $(P, \leq)$, also write P for the associated category. A P-persistence module is a functor over $\mathbb{F}$ is a functor $P \to \mathrm{Vect}_{\mathbb{F}}$. Maps between P-persistence modules are natural transformations.

We now follow the presentation in Section 4.5 of Carlsson–Vejdemo-Johansson[17] in order to classify $\mathbb{R}$-persistence modules.[18]

We begin with the free $\mathbb{R}$-persistence modules. Let X be a set and consider a function $\rho\colon X \to \mathbb{R}$. We may view (X, ρ) as an $\mathbb{R}$-filtered set via the sublevel filtration with $X_s := \rho^{-1}(-\infty, s] = \{x \in X \mid \rho(x) \leq s\}$. We write $\mathscr{V}(X, \rho)$ for the $\mathbb{R}$-persistence module with

$$\mathscr{V}(X,\rho)_s := \mathbb{F} \cdot X_s.$$

[17] Carlsson, G. and Vejdemo-Johansson, M. (2022). *Topological data analysis with applications*. Cambridge University Press, Cambridge

[18] Later, when we study multipersistence, we will see that for most posets P, P-persistence modules do not admit a classification (in the sense of being "wild type" problems in the language of representation theory).

Here $\mathbb{F} \cdot X_s$ is the $\mathbb{F}$-vector space with basis the set X_s; its elements are formal linear combinations of elements of X_s. The linear transformations $\iota_s^t \colon \mathscr{V}(X,\rho)_s \to \mathscr{V}(X,\rho)_t$ are the subspace inclusions induced by $X_s \subseteq X_t$.

Definition 3.2. We say that an $\mathbb{R}$-persistence module is *free* when it is isomorphic to an $\mathbb{R}$-persistence module of the form $\mathscr{V}(X,\rho)$; it is additionally *finitely generated* when X is a finite set.

Importantly, we can take quotients of $\mathbb{R}$-persistence modules. If $\mathscr{U} \leq \mathscr{V}$ is a sub-$\mathbb{R}$-persistence module (meaning $U_s \leq V_s$ for all s with compatible transition maps), then $\mathscr{V}/\mathscr{U}$ is the $\mathbb{R}$-persistence module with s-th vector space V_s/U_s and

$$\iota_s^t(v_s + U_s) = \iota_s^t(v_s) + U_t.$$

Given a map of a persistence modules $f \colon \mathscr{V} \to \mathscr{W}$, we define the *cokernel* of f to be

$$\operatorname{coker} f := \mathscr{W} / \operatorname{im} f$$

where $\operatorname{im} f$ is the sub-persistence module of $\mathscr{W}$ with $(\operatorname{im} f)_s = \operatorname{im} f_s$.

Definition 3.3. We say that an $\mathbb{R}$-persistence module is *finitely presented* when it is isomprhic to an $\mathbb{R}$-persistence module of the form $\operatorname{coker} f$ where f is a map between finitely generated free $\mathbb{R}$-persistence modules.

Recall from linear algebra that choosing bases for vector spaces V, W allows us to write down a matrix A_f[19] for every linear transformation $f \colon V \to W$ representing f. If the bases for V, W are $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, respectively, and $f(v_j) = \sum_i \lambda_i w_i$, then the v_j-column of A_f is $(\lambda_1, \ldots, \lambda_m)$. If U is another vector space with basis $u_1, \ldots, u_\ell$ and $g \colon U \to V$ is another linear transformation, then $A_{g \circ f} = A_g \cdot A_f$. This is the purpose of matrix multiplication.

[19] Importantly, A_f depends on the choice of bases, but we will not include the bases in our notation. Also note that we are being coy about the role that ordering of bases plays.

For a pair of finite sets X, Y, define an (X,Y)-matrix to be an array $[a_{xy}]_{(x,y)\in X\times Y}$ of elements of $\mathbb{F}$. Write $r(x)$ for the row associated with x, and $c(y)$ for the column associated with y. Given a finitely generated free persistence module $\mathscr{V}(X,\rho)$, note that

$$\mathscr{V}(X,\rho)_s = \mathbb{F} \cdot X_s = \mathbb{F} \cdot X \text{ for } s \gg 0$$

since X is finite. Thus any map $f \colon \mathscr{V}(Y,\sigma) \to \mathscr{V}(X,\rho)$ of finitely generated free persistence modules induces a linear transformation $f_\infty \colon \mathbb{F} \cdot Y \to \mathbb{F} \cdot X$ with an associated (X,Y)-matrix $[a_{xy}] := A_{f_\infty}$ where we work with the bases Y, X. This gives us an encoding of maps between finitely generated free $\mathbb{R}$-persistence modules:

Theorem 3.4. *Given f as above, the (X,Y)-matrix $A_{f_\infty} = [a_{xy}]$ satisfies $a_{xy} = 0$ whenever $\rho(x) > \sigma(y)$. Furthermore, any (X,Y)-matrix $A = [a_{xy}]$ satisfying this condition induces a map of $\mathbb{R}$-persistence modules $f_A \colon \mathscr{V}(Y,\sigma) \to \mathscr{V}(X,\rho)$ and the correspondences $f \mapsto A_{f_\infty}$ and $A \mapsto f_A$ are mutually inverse.*

Proof. First suppose $\rho(x) > \sigma(y)$. We have $f_\infty(y) = \sum_{x \in X} a_{xy} x$. Such a linear combination lies in $\mathscr{V}(X, \rho)_s$ if and only if $a_{xy} = 0$ for $\rho(x) > s$. Specializing to $s = \sigma(y)$ gives the first claim. We leave the second and third statements to the reader. □

Definition 3.5. Given $\mathbb{R}$-filtered finite sets (X, ρ) and (Y, σ), call an (X, Y)-matrix satisfying the condition of the theorem (ρ, σ)-*adapted*. Given a (ρ, σ)-adapted (X, Y)-matrix A, define

$$\theta(A) := \operatorname{coker} f_A.$$

We see straightaway that $\theta(A)$ is a finitely presented $\mathbb{R}$-persistence module, and the theorem implies that every finitely presented $\mathbb{R}$-persistence module is isomorphic to one of the form $\theta(A)$. Thus to classify finitely presented $\mathbb{R}$-persistence modules, it suffices to understand for which (X, Y)-matrices A, A' we have $\theta(A) \cong \theta(A')$.

Lemma 3.6. *Let (X, ρ) be a finite $\mathbb{R}$-filtered set. Then the automorphisms*[20] *of $\mathscr{V}(X, \rho)$ can be identified with the invertible (ρ, ρ)-adapted (X, X)-matrices.*[21] □

[20] An *automorphism* is an isomorphism with the same source and target.

[21] Unpacking this, we have an invertible (X, X)-matrix $[a_{xx'}]_{(x,x') \in X \times X}$ satisfying $a_{xx'} = 0$ for $\rho(x) > \rho(x')$. If X is ordered by increasing ρ values, this is an invertible upper triangular matrix, so upper triangular with no 0's on the diagonal.

The following proposition now follows by abstract nonsense:

Proposition 3.7. *Let (X, ρ) and (Y, σ) be finite $\mathbb{R}$-filtered sets, and let A be a (ρ, σ)-adapted (X, Y)-matrix. Let B be an invertible (ρ, ρ)-adapted (X, X)-matrix and let C be an invertible (σ, σ)-adapted (Y, Y)-matrix. Then BAC is a (ρ, σ)-adapted (X, Y)-matrix and*

$$\theta(A) \cong \theta(BAC).$$

We now establish notation that will allows us to state our classification theorem. For $a \in \mathbb{R}$ and $b \in \mathbb{R} \cup \infty$ with $a < b$, define

$$\mathbb{I}[a, b)_s := \begin{cases} \mathbb{F} & \text{if } a \leq s < b \\ 0 & \text{otherwise} \end{cases}$$

and $\iota_s^t \colon \mathbb{I}[a, b)_s \to \mathbb{I}[a, b)_t$ to be $\mathrm{id}_{\mathbb{F}}$ if $a \leq s \leq t < b$ and 0 otherwise. This is the *interval $\mathbb{R}$-persistence module* associated with $[a, b)$.

Lemma 3.8. *The interval $\mathbb{R}$-persistence module $\mathbb{I}[a, b)$ is finitely presented.*

Proof. For $b < \infty$, let $(X, \rho) = (\{x\}, \rho(x) = a)$ and $(Y, \sigma) = (\{y\}, \sigma(y) = b)$. One may check that the 1×1 matrix $[1]$ is (ρ, σ)-adapted with $\mathbb{I}[a, b) = \theta([1])$. If $b = \infty$, then $\mathbb{I}[a, \infty) = \mathscr{V}(\{x\}, \rho(x) = a) = \theta(\varnothing)$ where $\varnothing$ is the 1×0 matrix representing the map $0 \to V(\{x\}, \rho(x) = a)$. □

This brings us to our classification theorem for finitely presented $\mathbb{R}$-persistence modules over $\mathbb{F}$:

Theorem 3.9. *Every finitely presented $\mathbb{R}$-persistence module over $\mathbb{F}$ is isomorphic to a finite direct sum of the form*

$$\mathbb{I}[a_1, b_1) \oplus \mathbb{I}[a_2, b_2) \oplus \cdots \oplus \mathbb{I}[a_n, b_b)$$

for some $a_i \in \mathbb{R}$, $b_i \in \mathbb{R} \cup \infty$, and $a_i < b_i$ for all i. Moreover,

$$\bigoplus_{i \in I} \mathbb{I}[a_i, b_i) \cong \bigoplus_{j \in J} \mathbb{I}[c_j, d_j)$$

for I, J finite sets if and only if $|I| = |J|$ and the multiset[22] *of intervals $[a_i, b_i)$ equals the multiset of intervals $[c_j, d_j)$.*

[22] A *multiset* is a set "with multiplicity". This can be formalized as a set X together with a function $m\colon X \to \mathbb{N}$ counting multiplicity.

Proof. First note that if A is the (ρ, σ)-adapted (X, Y) matrix with 1's in positions $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and 0's elsewhere and there is at most one 1 in every row and column, then

$$\theta(A) \cong \bigoplus_{i=1}^{n} \mathbb{I}[\rho(x_i), \sigma(y_i)) \oplus \bigoplus_{x \in X \smallsetminus \{x_1, \ldots, x_n\}} \mathbb{I}[\rho(x), \infty).$$

Thus given an arbitrary (ρ, σ)-adapted (X, Y)-matrix A, it suffices to construct an invertible (ρ, ρ)-adapted (X, X)-matrix B and invertible (σ, σ)-adapted (Y, Y)-matrix C such that every row and column of BAC has at most one 1 with all other entries 0.

We accompish this task via (ρ, σ)*-adapted row and column operations*. These are scalings of rows or columns by a nonzero element of $\mathbb{F}$, adding a multiple of $r(x)$ to $r(x')$ when $\rho(x) \geq \rho(x')$, and adding a multiple of $c(y)$ to $c(y')$ when $\sigma(y) > \sigma(y')$. The reader should check that these may be accomplished by multiplying A on the left or right by the appropriate kind of invertible matrix.

Now find y maximizing $\sigma(y)$ over all y with $c(y) \neq 0$. Then find x maximizing $\rho(x)$ over x such that $a_{xy} \neq 0$. We are free to add multiples of $r(x)$ to all other rows to cancel out $c(y)$ except for a_{xy}. We are then further free to add multiples of $c(y)$ to cancel out $r(x)$ except for a_{xy}. Multiplying $r(x)$ by $1/a_{xy}$ we get 1 in the xy-position. Now keep repeating the process with the next largest $\sigma(y)$ and $\rho(x)$ with $a_{xy} \neq 0$, and we eventually get the desired form.

The example following this proof implements the algorithm of this paragraph. The reader may wish to read the example in parallel with the proof.

We leave the uniqueness statement to the reader, who can also look at Proposition 4.52 of Carlsson–Vejdemo-Johansson. □

Example 3.10. Let us demonstrate the (ρ, σ)-adapted Gaussian elimination algorithm from the above proof. Let $X = \{x_1, \ldots, x_6\}$ with

$$\rho(x_1) = \rho(x_2) = 0, \quad \rho(x_3) = \rho(x_4) = \rho(x_5) = 1, \quad \rho(x_6) = 2$$

and $Y = \{y_1, \ldots, y_5\}$ with

$$\sigma(y_1) = 1, \quad \sigma(y_2) = \sigma(y_3) = 2, \quad \sigma(y_4) = \sigma(y_5) = 3.$$

Using the natural orderings of X and Y, we see that

$$A = \begin{bmatrix} 0 & 4 & 16 & 12 & 4 \\ 6 & 14 & 2 & 0 & 0 \\ 0 & 0 & 2 & 4 & 0 \\ 0 & 2 & 0 & 2 & 2 \\ 0 & 2 & 4 & 2 & 4 \\ 0 & 1 & 2 & 1 & 2 \end{bmatrix}$$

is (ρ, σ)-adapted. (In fact, the only condition is that the (x_6, y_1)-entry in the bottom left is 0.) We begin with the algorithm with the bottom right (x_6, y_5)-entry and use the bottom row to cancel out the rest of the rightmost column. This results in

$$\begin{bmatrix} 0 & 2 & 12 & 10 & 0 \\ 6 & 14 & 2 & 0 & 0 \\ 0 & 0 & 2 & 4 & 0 \\ 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \end{bmatrix}.$$

We then clear out the bottom row with the rightmost column and multiply the bottom row by $1/2$ to get

$$\begin{bmatrix} 0 & 2 & 12 & 10 & 0 \\ 6 & 14 & 2 & 0 & 0 \\ 0 & 0 & 2 & 4 & 0 \\ 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We now work at the (x_3, y_4)-entry and clear out the rest of the y_4-column by adding $-5/2r(x_3)$ to $r(x_1)$ to get

$$\begin{bmatrix} 0 & 2 & 7 & 0 & 0 \\ 6 & 14 & 2 & 0 & 0 \\ 0 & 0 & 2 & 4 & 0 \\ 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now using $c(y_4)$ to zero out the rest of $r(x_3)$ and then multiplying $r(x_3)$ by

$1/4$ we get

$$\begin{bmatrix} 0 & 2 & 7 & 0 & 0 \\ 6 & 14 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Moving on to the (x_4, y_3)-entry the next round of row and column operations produces

$$\begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 6 & 14 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now working in position (x_2, y_2) we finally arrive at

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

This implies that the interval decomposition of $\theta(A)$ is

$$\begin{aligned} \theta(A) &\cong \mathbb{I}[\rho(x_1), \infty) \oplus \mathbb{I}[\rho(x_2), \sigma(y_2)) \oplus \mathbb{I}[\rho(x_3), \sigma(y_4)) \oplus \mathbb{I}[\rho(x_4), \sigma(y_3)) \oplus \mathbb{I}[\rho(x_5), \infty) \oplus \mathbb{I}[\rho(x_6), \sigma(y_5)) \\ &= \mathbb{I}[0, \infty) \oplus \mathbb{I}[0, 2) \oplus \mathbb{I}[1, 3) \oplus \mathbb{I}[1, 2) \oplus \mathbb{I}[1, \infty) \oplus \mathbb{I}[2, 3) \end{aligned}$$

As described in the first lecture, we will record the isomorphism type of an $\mathbb{R}$-persistence module with a *barcode* or *persistence diagram*. The barcode of

$$\mathbb{I}[a_1, b_1) \oplus \mathbb{I}[a_2, b_2) \oplus \cdots \oplus \mathbb{I}[a_n, b_b)$$

is the multiset of intervals $[a_i, b_i)$, typically drawn as stacked horizontal intervals. The persistence diagram of the same $\mathbb{R}$-persistence module is the multiset of points (a_i, b_i) in the extended plane $\mathbb{R} \times (\mathbb{R} \cup \infty)$. Since $a_i \leq b_i$, all these points lie on or above the line of slope 1 through the origin. The persistence diagram for the above example is displayed in the margin.

3.1 Notes

In the first lecture, we classified tame $\mathbb{N}$-persistence modules via the theory of finitely generated graded modules over a graded PID. Here we classified finitely presented $\mathbb{R}$-persistence modules essentially through a modified Gaussian elimination algorithm. Gaussian elimination requires $O(n^3)$ arithmetic operations, so the algorithm implicit in our proof is polynomial time but still computationally expensive. Zomorodian[23] provides methods for speeding up these algorithms when working with the persistent homology of particular types of simplicial complexes — called tidy sets — that include Vietoris–Rips complexes of point cloud data. This is one reason for going beyond Čech complexes when considering filtered spaces induced by point clouds, as we shall in the next lecture.

[23] Zomorodian, A. (2010). The tidy set: a minimal simplicial set for computing homology of clique complexes. In *Proceedings of the Twenty-Sixth Annual Symposium on Computational Geometry*, SoCG '10, page 257âĂŞ266, New York, NY, USA. Association for Computing Machinery

3.2 *Exercises*

(1) Let $\mathbb{F} = \mathbb{Q}$, $X = \{0,2,4\}$, $Y = \{1,2,3,4\}$, $\rho(x) = x$, and $\sigma(y) = y$. Check that

$$A = \begin{bmatrix} 2 & 1 & 2 & 0 \\ 0 & 2 & 4 & 3 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

is a (ρ, σ)-adapted (X,Y)-matrix (with the natural order of rows and columns) and then implement the algorithm from the proof of Theorem 3.9 to determine the persistence diagram associated with $\theta(A)$.

(2) Your answer to the previous question should have a 1 in the $(4,4)$-position (so third row, fourth column), so one of the summands of the interval decomposition of $\theta(A)$ is $\mathbb{I}[4,4)$. How do we interpret this $\mathbb{R}$-persistence module?

(3) Suppose A is a (ρ, σ)-adapted (X,Y)-matrix and let $x_{\max}$ and $y_{\max}$ be elements of X and Y maximizing ρ and σ, respectively. Assume that the $y_{\max}$ column of A is nonzero. Prove that the interval decomposition of $\theta(A)$ includes $\mathbb{I}[\rho(x_{\max}), \sigma(y_{\max}))$.

(4) Suppose $|X| = |Y|$ and the determinant of a (ρ, σ)-adapted matrix A is nonzero. What can you say about the associated interval decomposition of $\theta(A)$?

4 Čech and Vietoris–Rips filtered complexes — 3 April

Today we will analyze the two main filtered spaces (abstract simplicial complexes) that arise from point clouds, the Čech and Vietoris–Rips filtered complexes. In preparation, we will discuss the nerve construction and lemma, a crucial tool guaranteeing that, under certain hypotheses, the Čech complex recovers the homotopy type of a space from which points are sampled. We will also recast point clouds as finite metric spaces, a framework that will allow us to apply persistence homology to additional data types.

4.1 The nerve of an open cover

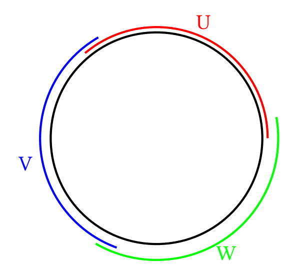
Let X be a topological space. An *open cover* of X is a collection $\mathfrak{U} = \{U_\alpha \mid \alpha \in A\}$ of open sets of X such that

$$X = \bigcup_{\alpha \in A} U_\alpha.$$

We have followed tradition in using mathfrak U, $\mathfrak{U}$, for an open cover. When we need another open cover we'll use mathfrak V, *i.e.*, $\mathfrak{V}$. We leave the reason for the incscrutability of this font as an exercise for the reader.

Given an open cover $\mathfrak{U}$, its *nerve* $N(\mathfrak{U})$ is the abstract simplicial complex with vertex set $\mathfrak{U}$ and a subset $\{U_\beta \mid \beta \in B \subseteq A\}$ in $N(\mathfrak{U})$ if and only if

$$\bigcap_{\beta \in B} U_\beta \neq \varnothing.$$



Example 4.1. Suppose $X = S^1$ and $\mathfrak{U} = \{U, V, W\}$ as illustrated. Then the vertices of $N(\mathfrak{U})$ are U, V, W (now viewed as 0-dimensional points), the edges of $N(\mathfrak{U})$ are $\{U, V\}$, $\{V, W\}$, and $\{W, U\}$. There are no 2- or higher-dimensional simplices as $U \cap V \cap W = \varnothing$.

The following Nerve Lemma tells us that nice open covers of nice spaces have nerves homotopy equivalent to the original space.

Lemma 4.2 (Nerve Lemma)**.** *Suppose $\mathfrak{U}$ is an open cover of a paracompact*[24] *space X. Suppose further that every nonempty intersection of finitely many elements of $\mathfrak{U}$ is contractible. Then*

$$X \simeq N(\mathfrak{U}).$$

Proof. This is Corollary 4G.3 of Hatcher.[25] The entirety of Section 4G is a fun read which is secretly about homotopy colimits. □

[24] A space is *paracompact* when every open cover has an open refinement that is locally finite. Here $\mathfrak{V}$ is a *refinement* of $\mathfrak{U}$ when every set in $\mathfrak{V}$ is a subset of a set in $\mathfrak{U}$. A cover is *locally finite* when every point in X has an open neighborhood that intersects only finitely many sets in the cover. Note that every subspace of $\mathbb{R}^N$ and every metric space is paracompact.

[25] Hatcher, A. (2002). *Algebraic topology*. Cambridge University Press, Cambridge

4.2 The Čech filtered complex

We now recast[26] the Čech filtered complex from the first lecture in terms of nerves. Fix a point cloud $P \subseteq \mathbb{R}^N$ and a scale $s \in \mathbb{R}$. Let $\mathfrak{U}_s(P) := \{B_r(x) \mid x \in P\}$ be the collection of radius r open balls centered at points in P; this is (tautologically) a cover of the subspace

$$X_s := \bigcup_{x \in P} B_r(x) \subseteq \mathbb{R}^N.$$

[26] In the first lecture, I used closed balls instead of open balls to define the Čech complex. We are now switching to the more standard convention of using open balls.

The *Čech complex* of P at scale s is the abstract simplicial complex

$$\check{\mathbb{C}}_s(P) := N(\mathfrak{U}_s(P)).$$

By labeling the vertices of $\check{\mathbb{C}}_s(P)$ by elements of P (rather than by open balls $B_r(x)$ for $x \in P$), we see that there is a natural simplicial inclusion $\check{\mathbb{C}}_s(P) \subseteq \check{\mathbb{C}}_t(P)$ for $s \leq t$. This makes

$$\check{\mathbb{C}}(P) := \{\check{\mathbb{C}}_s(P) \mid s \in \mathbb{R}\}$$

into an $\mathbb{R}$-filtered abstract simplicial complex which we call the *Čech filtered complex* of P.

4.3 *The Vietoris–Rips filtered complex*

As demonstrated by the Nerve Lemma (Lemma 4.2), the Čech construction has very desirable theoretical properties. It is not, though, always the best tool for the job. First, it is computationally expensive to determine when k-fold intersections of open balls are nonempty. Second, our "data" might not come from a point cloud in $\mathbb{R}^N$ or indeed be sampled from any ambient metric space. For these reasons, we introduce the Vietoris–Rips complex of a (usually finite) metric space P.

Recall that a *metric space* is a set P equipped with a function $d\colon P \times P \to \mathbb{R}_{\geq 0}$ such that

» $d(x,y) = 0$ if and only if $x = y$ (identity of indiscernibles),

» $d(x,y) = d(y,x)$ (symmetry), and

» $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality).

When P is a point cloud in $\mathbb{R}^N$, we may endow it with the *induced metric*, where $d(x,y) = \|y - x\|$, but there are many other examples of metric spaces, many of which do not embed isometrically in Euclidean space. (See Morgan[27] for precise criteria on when this is possible.)

[27] Morgan, C. L. (1974). Embedding matric spaces in Euclidean space. *J. Geom.*, 5:101–107

Definition 4.3. The *Vietoris–Rips complex* of a metric space (P,d) at scale $s \in \mathbb{R}$, denoted $\mathbb{VR}_s(P)$, is the abstract simplicial complex with vertex set P and for which $\{x_0, \ldots, x_k\} \subseteq P$ is a k-simplex in $\mathbb{VR}_s(P)$ if and only if

$$d(x_i, x_j) < s$$

for $0 \leq i < j \leq k$.

It is clear that $\mathbb{VR}_s(P) \subseteq \mathbb{VR}_t(P)$ for $s \leq t$. Further, we see that $\{x_0, \ldots, x_k\}$ is in $\mathbb{VR}_s(P)$ if and only if each $\{x_i, x_j\}$ is in $\mathbb{VR}_s(P)$. This makes $\mathbb{VR}_s(P)$ a *flag complex* — a simplicial complex determined by its 1-skeleton.

Slightly more subtle is the relation between Vietoris–Rips and Čech complexes:

Proposition 4.4. *For a point cloud $P \subseteq \mathbb{R}^N$ and scale $s \geq 0$, we have*

$$\mathbb{VR}_{s/2}(P) \subseteq \check{\mathbb{C}}_s(P) \subseteq \mathbb{VR}_{2s}(P).$$

Furthermore, the flag complex on $\check{\mathbb{C}}_s(P)$ is exactly $\mathbb{VR}_{2s}(P)$.

Every abstract simplicial complex K induces a flag complex $\mathscr{F}K$ determined by the 1-skeleton of K. The set $\{x_0, \ldots, x_k\}$ is a k-simplex of $\mathscr{F}K$ if and only if each $\{x_i, x_j\}$ is a 1-simplex of K. Of course, we always have $K \subseteq \mathscr{F}K$.

Proof. We begin by checking that $\check{\mathbb{C}}_s(P) \subseteq \mathbb{VR}_{2s}(P)$. We must verify that $\{x_0, \ldots, x_k\} \in \check{\mathbb{C}}_s(P)$ implies $d(x_i, x_j) < 2s$ for all i, j. Since the intersection of all the $B_s(x_i)$ is nonempty, we may choose some z in their intersection. By the triangle inequality,

$$d(x_i, x_j) \leq d(x_i, z) + d(z, x_j) < 2s,$$

as desired.

Next, suppose that $\{x_0, \ldots, x_k\}$ is a k-simplex of $\mathbb{VR}_{s/2}(P)$. Since $d(x_0, x_1) < s/2$, we may choose $z \in B_{s/4}(x_0) \cap B_{s/4}(x_1)$, in which case $d(z, x_0) < s/4$. For $0 \leq i \leq k$ we have $d(z, x_i) \leq d(z, x_0) + d(x_0, x_i) < 3s/4 < s$. Thus $z \in \bigcap_{i=0}^k B_s(x_i) \neq \varnothing$, so $\{x_0, \ldots, x_k\}$ is a k-simplex of $\check{\mathbb{C}}_s(P)$. This shows that $\mathbb{VR}_{s/2}(P) \subseteq \check{\mathbb{C}}_s(P)$.

For the final assertion, we already know that $\mathbb{VR}_{2s}(P)$ is flag and contains $\check{\mathbb{C}}_s(P)$, so we automatically have $\mathscr{F}\check{\mathbb{C}}_s(P) \subseteq \mathbb{VR}_{2s}(P)$. It remains to show that if $\{x_0, \ldots, x_k\}$ is a k-simplex of $\mathbb{VR}_{2s}(P)$, then each $\{x_i, x_j\}$ is a 1-simplex of $\check{\mathbb{C}}_s(P)$. We know that $d(x_i, x_j) < 2s$, so the midpoint of the line segment connecting x_i and x_j witnesses that $B_s(x_i) \cap B_s(x_j) \neq \varnothing$. This completes the argument. $\square$

5 Distance and stability for persistent homology — 8 April

References

Bradley, T.-D., Bryson, T., and Terilla, J. (2020). *Topology—a categorical approach*. MIT Press, Cambridge, MA.

Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308.

Carlsson, G. and Vejdemo-Johansson, M. (2022). *Topological data analysis with applications*. Cambridge University Press, Cambridge.

Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75.

Hatcher, A. (2002). *Algebraic topology*. Cambridge University Press, Cambridge.

Morgan, C. L. (1974). Embedding matric spaces in Euclidean space. *J. Geom.*, 5:101–107.

Munkres, J. R. (2000). *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, second edition.

Oudot, S. Y. (2015). *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Tralie, C., Saul, N., and Bar-On, R. (2018). Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925.

Virk, Ž. (2022). Introduction to persistent homology. `https://zalozba.fri.uni-lj.si/virk2022.pdf`. Accessed on 19 March 2024.

Zomorodian, A. (2010). The tidy set: a minimal simplicial set for computing homology of clique complexes. In *Proceedings of the Twenty-Sixth Annual Symposium on Computational Geometry*, SoCG '10, page 257âĂŞ266, New York, NY, USA. Association for Computing Machinery.