Topological Persistence

Course Notes

Steve Oudot

In the following, we consider simplicial homology with coefficients in a field $\boldsymbol{k}$, omitted in the notations. By $\mathsf{H}_p(X)$ we denote the p-th homology group of space X. For background on homology theory, see Chapter 1 of [13].

1 Persistence modules

Definition 1. Given an index set $T \subseteq \mathbb{R}$, a *filtration* over T is a family of topological spaces $\mathcal{F} = \{F_t\}_{t\in T}$ that is nested with respect to inclusion, that is:

$$\forall t \leq t' \in T, \quad F_t \subseteq F_{t'}.$$

Examples of such filtrations include:

- the offsets $P^t = \bigcup_{p\in p} B(p,t)$ of a compact set $P \subset \mathbb{R}^d$, where $B(p,r)$ is a shorthand for the Euclidean ball of center p and radius t;
- the sublevel sets $F_t = f^{-1}((-\infty, t])$ or superlevel sets $F^t = f^{-1}([t, +\infty))$ of a real-valued map.

Applying the p-th homology functor H_p to a filtration, we obtain a persistence module, i.e. a family of vector spaces and linear maps as follows:

Definition 2. Given an index set $T \subseteq \mathbb{R}$, a *persistence module* $\mathbb{V}$ over T is a family of vector spaces $\{V_t\}_{t\in T}$ together with a family of linear maps $\{v_t^{t'} : V_t \to V_{t'}\}_{t\leq t'\in T}$ that satisfies the following conditions (where $\mathbb{1}_X$ denotes the identity of X):

$$\forall t \in T, \quad v_t^t = \mathbb{1}_{V_t},$$
$$\forall t \leq t' \leq t'' \in T, \quad v_{t'}^{t''} \circ v_t^{t'} = v_t^{t''}.$$

Let $\mathbb{V}$ be a persistence module over an index set $T \subseteq \mathbb{R}$. We want to decompose $\mathbb{V}$ as a direct sum of indecomposable persistence modules:

$$\mathbb{V} \cong \bigoplus_{j\in J} \mathbb{V}^j, \tag{1}$$

where each indecomposable summand $\mathbb{V}^j$ is an *interval module* $\mathbb{I}[b_j^*, d_j^*]$, defined as follows:

$$\mathbb{I}[b_j^*, d_j^*] \quad = \quad \underbrace{0 \xrightarrow{0} \cdots \xrightarrow{0} 0}_{i<b_j^*} \xrightarrow{0} \underbrace{\boldsymbol{k} \xrightarrow{\mathbb{1}} \cdots \xrightarrow{\mathbb{1}} \boldsymbol{k}}_{[b_j^*,\, d_j^*]} \xrightarrow{0} \underbrace{0 \xrightarrow{0} \cdots \xrightarrow{0} 0}_{i>d_j^*} \tag{2}$$

Note that the direct sum in (1) is defined *pointwise*, that is, every space V_t is isomorphic to the direct sum of the spaces V_t^j. The notation $[b_j^*, d_j^*]$ in (2) indicates that the interval may be indifferently closed, open, or half-open.

Theorem 3 (Decomposition)**.** *A decomposition such as (1) holds at least in the following cases:*

- *when the index set T is finite and the spaces V_t of $\mathbb{V}$ are arbitrary [1, 10],*
- *when T is arbitrary and every space V_t is finite-dimensional ($\mathbb{V}$ is then called* pointwise finite-dimensional*) [7, 14].*

Moreover, whenever it exists, the decomposition is independent of the choice of base field $\boldsymbol{k}$, and it is unique up to isomorphism and permutation of the terms in the direct sum [2].

When a persistence module $\mathbb{V}$ decomposes into interval modules as in (1), its algebraic structure is fully described by the corresponding collection of intervals, called its *barcode*. (Almost) equivalently, the decomposition can be represented as a *persistence diagram*, i.e. a multiset of points in the extended plane $\bar{\mathbb{R}}^2 = [-\infty, +\infty]^2$, where each point (b, d) corresponds to a copy of the interval $[b^*, d^*]$:

$$\mathsf{dgm}(\mathbb{V}) = \{(b_j, d_j) \mid j \in J\} \subset \bar{\mathbb{R}}^2.$$

Unfortunately, not all persistence modules of interest in applications can be decomposed into interval modules as in (1). Here is an example:

Example 4. Take the compact set $\{0\} \cup \{(\frac{1}{n}, 0) \mid n \geq 1\}$ in the plane, and consider its offsets filtration, indexed over the infinite set $T = \mathbb{R}^+$. The induced persistence module in 0-th homology is also indexed over T, and at $t = 0$ it has infinite dimension, therefore it satisfies none of the sufficient conditions from Theorem 3.

The solution to this is the concept of $\mathfrak{q}$-*tameness*, defined as follows:

Definition 5. A persistence module $\mathbb{V}$ indexed over T is $\mathfrak{q}$-*tame* if the maps $v_t^{t'}$ have finite rank for all $t < t' \in T$. Note the difference with the pointwise finite-dimensionality from Theorem 3, in which the maps $v_t^t = \mathbb{1}_{V_t}$ also have finite rank.

Even though they may not always be decomposable into intervals, $\mathfrak{q}$-tame modules can be assigned well-defined persistence diagrams [3, 5]:

Theorem 6. *Every $\mathfrak{q}$-tame persistence module $\mathbb{V}$ has a well-defined persistence diagram $\mathsf{dgm}(\mathbb{V})$.*

2 Stability

The space of persistence diagrams is naturally equipped with a matching distance called the *bottleneck distance*. Specifically, a partial matching M between two multisets A, B in the plane, denoted $M : A \leftrightarrow B$, is a subset of the product $A \times B$ whose canonical projections onto A and B are injective, that is,

$$\begin{aligned} \forall a \in A, &\ \text{ there is at most one } b \in B \text{ s.t. } (a, b) \in M, \\ \forall b \in B, &\ \text{ there is at most one } a \in A \text{ s.t. } (a, b) \in M. \end{aligned}$$

The cost of a matched pair (a, b) is their ℓ_∞-distance $\|a - b\|_\infty$, while the cost of an unmatched point $x \in A \cup B$ is its ℓ_∞-distance to the diagonal $\Delta = \{(y, y) \mid y \in \bar{\mathbb{R}}\}$. The bottleneck cost of M, denoted $c(M)$, is then the supremum of these costs, and the bottleneck distance between A and B, denoted $\mathrm{d_b}(A, B)$, is the infimum of the bottleneck costs over all partial matchings $A \leftrightarrow B$:

$$\mathrm{d_b}(A, B) = \inf_{M : A \leftrightarrow B} c(M). \tag{3}$$

Our first stability result concerns persistence modules induced at the homology level by sublevel-sets filtrations of functions. It was proven in a restricted form in [6], then in its general form in [5]. We call q-*tame* any real-valued function whose sublevel-sets filtration induces a q-tame persistence module at the p-th homology level for every homology dimension p. In this case we denote by $\mathsf{dgm}_p(f)$ the associated persistence diagram, and by $\mathsf{dgm}(f)$ the collection of these diagrams for all $p \in \mathbb{N}$. Given another real-valued function g, the bottleneck distance between their persistence diagrams is computed for each homology dimension p separately, and the supremum over all $p \in \mathbb{N}$ is denoted by $\mathrm{d_b}(\mathsf{dgm}(f), \mathsf{dgm}(g))$.

Theorem 7 (Stability for functions)**.**
For any q-*tame functions* $f, g : X \to \mathbb{R}$*,* $\mathrm{d_b}(\mathsf{dgm}(f), \mathsf{dgm}(g)) \leq \|f - g\|_\infty$.

The proof of this theorem relies on the observation that, when $\|f - g\|_\infty$ is at most say ε, the sublevel-sets filtrations $\mathcal{F}$ of f and $\mathcal{G}$ of g are ε-*interleaved* as follows:

$$\forall t \in \mathbb{R},\ F_t \subseteq G_{t+\varepsilon} \text{ and } G_t \subseteq F_{t+\varepsilon}. \tag{4}$$

More precisely, for all indices $t \leq t' \in \mathbb{R}$ we have the following commutative diagram of sublevel sets and inclusion maps:

$$\begin{tikzcd}
F_t \arrow[rr] \arrow[dr] & & F_{t'} \arrow[dr] & \\
& G_{t+\varepsilon} \arrow[rr] & & G_{t'+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& F_{t+\varepsilon} \arrow[rr] & & F_{t'+\varepsilon} \\
G_t \arrow[ur] \arrow[rr] & & G_{t'} \arrow[ur] &
\end{tikzcd}$$

$$\begin{tikzcd}
F_t \arrow[rr] \arrow[dr] & & F_{t+2\varepsilon} \\
& G_{t+\varepsilon} \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& F_{t+\varepsilon} \arrow[dr] & \\
G_t \arrow[ur] \arrow[rr] & & G_{t+2\varepsilon}
\end{tikzcd}$$

After applying the p-th homology functor H_p, we get a commutative diagram of vector spaces and linear maps, which involves the persistence modules $\mathsf{H}_p(\mathcal{F})$ and $\mathsf{H}_p(\mathcal{G})$ together with the cross maps $\phi_t : \mathsf{H}_p(F_t) \to \mathsf{H}_p(G_{t+\varepsilon})$ and $\psi_t : \mathsf{H}_p(G_t) \to \mathsf{H}_p(F_{t+\varepsilon})$ induced at the homology level by the inclusions $F_t \hookrightarrow G_{t+\varepsilon}$ and $G_t \hookrightarrow F_{t+\varepsilon}$. For simplicity of notations, let us rename $\mathsf{H}_p(\mathcal{F}) = \mathbb{V}$ and $\mathsf{H}_p(\mathcal{G}) = \mathbb{W}$:

$$\begin{tikzcd}
V_t \arrow[rr] \arrow[dr, "\phi_t"] & & V_{t'} \arrow[dr, "\phi_{t'}"] & \\
& W_{t+\varepsilon} \arrow[rr] & & W_{t'+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& V_{t+\varepsilon} \arrow[rr] & & V_{t'+\varepsilon} \\
W_t \arrow[ur, "\psi_t"] \arrow[rr] & & W_{t'} \arrow[ur, "\psi_{t'}"] &
\end{tikzcd} \tag{5}$$

$$\begin{tikzcd}
V_t \arrow[rr] \arrow[dr, "\phi_t"'] & & V_{t+2\varepsilon} \\
& W_{t+\varepsilon} \arrow[ur, "\psi_{t+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& V_{t+\varepsilon} \arrow[dr, "\phi_{t+\varepsilon}"] & \\
W_t \arrow[ur, "\psi_t"] \arrow[rr] & & W_{t+2\varepsilon}
\end{tikzcd} \tag{6}$$

This is what we call an ε-*interleaving* between persistence modules. As one can see, it is the direct translation, at the algebraic level, of the interleaving between filtrations, although it does not actually need filtrations to start with in order to be stated.

Definition 8. Let $\mathbb{V}, \mathbb{W}$ be two persistence modules over $\mathbb{R}$, and let $\varepsilon \geq 0$. An ε-*interleaving* between $\mathbb{V}, \mathbb{W}$ is given by two families of linear maps $(\phi_t : V_t \to W_{t+\varepsilon})_{t \in \mathbb{R}}$ and $(\psi_t : W_t \to V_{t+\varepsilon})_{t \in \mathbb{R}}$

such that the diagrams (5) and (6) commute for all $t \leq t' \in \mathbb{R}$. The *interleaving distance* between $\mathbb{V}$ and $\mathbb{W}$ is

$$\mathrm{d_i}(\mathbb{V}, \mathbb{W}) = \inf \{\varepsilon \geq 0 \mid \text{there is an } \varepsilon\text{-interleaving between } \mathbb{V} \text{ and } \mathbb{W}\}. \tag{7}$$

Note that there are no conditions on the persistence modules $\mathbb{V}, \mathbb{W}$, which can be arbitrary as long as they are defined over the same ground field $\boldsymbol{k}$. When there is no ε-interleaving between $\mathbb{V}$ and $\mathbb{W}$ for any $\varepsilon \geq 0$, we let $\mathrm{d_i}(\mathbb{V}, \mathbb{W}) = +\infty$.

We can now rephrase Theorem 7 at the algebraic level directly. The proof can be found in [5].

Theorem 9 (Stability for persistence modules)**.**
For any q*-tame persistence modules* $\mathbb{V}, \mathbb{W}$*,* $\mathrm{d_b}(\mathsf{dgm}(\mathbb{V}), \mathsf{dgm}(\mathbb{W})) \leq \mathrm{d_i}(\mathbb{V}, \mathbb{W})$.

It turns out that the inequality is in fact an equality, so the map $\mathbb{V} \mapsto \mathsf{dgm}(\mathbb{V})$ is an isometry from the space of q-tame persistence modules into the space of persistence diagrams. The converse inequality $\mathrm{d_b}(\mathsf{dgm}(\mathbb{V}), \mathsf{dgm}(\mathbb{W})) \geq \mathrm{d_i}(\mathbb{V}, \mathbb{W})$ is proven in [11].

3 Computation

From now on we assume for simplicity that the field of coefficients is $\boldsymbol{k} = \mathbb{Z}_2$. The persistence algorithm can be adapted to work with any arbitrary finite field. Suppose we are given a filtration $\mathcal{K}$ composed of finitely many finite simplicial complexes:

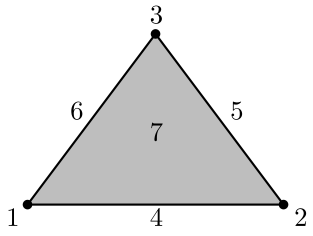
$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m = K. \tag{8}$$

Suppose also that the difference $K_i \setminus K_{i-1}$ is composed of only one simplex σ_i for every $i = 1, \cdots, m$. Thus, the filtration is obtained by adding one simplex at a time, and the insertion order is compatible with the face order in K, that is, every simplex appears after its proper faces.

Given this input, the algorithm builds a square $m \times m$ matrix M representing the boundary operator on the chains of simplices of K of all dimensions. Specifically, M has one row and one column per simplex of K, with $M_{ij} = 1$ if σ_i is a face of codimension 1 of σ_j and $M_{ij} = 0$ otherwise. Moreover, for the needs of the algorithm, the rows and columns of M are ordered as the simplices in the sequence of (8). Since the sequence is compatible with the face order in K, the matrix M is upper-triangular.

Once M is built, the algorithm adds columns from left to right until the matrix is reduced to column-echelon form. The pseudo-code is given in Algorithm 1, where $\mathsf{low}(j, R)$ denotes the row index of the lowest nonzero entry in column j of the matrix R—$\mathsf{low}(j) = 0$ if column j is entirely zero.

```
Algorithm 1: Matrix reduction
Input: $m \times m$ binary matrix $M$
Let $R = M$;
for $j = 1$ to $m$ do
    while there exists $k < j$ with $\mathsf{low}(k, R) = \mathsf{low}(j, R) \neq 0$ do
        add (modulo 2) column $k$ to column $j$ in $R$;
    end
end
Output: $R$
```



	1	2	3	4	5	6	7
1				*		*	
2				*	*		
3					*	*	
4							*
5							*
6							*
7							

	1	2	3	4	5	6	7
1				*			
2				1	*		
3					1		
4							*
5							*
6							1
7							

Figure 1: Left: filtration of a solid triangle (each simplex is marked with its index in the sequence). Center: the corresponding boundary matrix M, where stars mark the nonzero entries. Right: the reduced matrix R, where lowest nonzero entries are marked with 1's while the other nonzero entries are marked with stars.

Upon termination, the matrix R has the desired column-echelon form. In particular, it has the property that the lowest nonzero entry of every nonzero column lies in a unique row. Its structure is interpreted as follows:

- every zero (resp. nonzero) column j corresponds to a positive (resp. negative) simplex σ_j,
- every nonzero column j is paired with the column $i = \mathsf{low}(j, R)$ and gives rise to a summand $\mathbb{I}[i, j)$ in the interval decomposition of the induced persistence module $\mathsf{H}_p(\mathcal{K})$ at the p-th homology level, where $p = \dim \sigma_i$,
- every unpaired zero column j gives rise to a summand $\mathbb{I}[j, +\infty)$ in the interval decomposition of $H_p(\mathcal{K})$, where $p = \dim \sigma_j$.

The crux of the analysis it to prove this interpretation correct. Formally:

Theorem 10 ([8])**.** *Unpon termination, the simplicial chains $\hat{\sigma}_1, \cdots, \hat{\sigma}_n$ represented by the columns of R yield a partition $F \sqcup G \sqcup H$ of the index set $\{1, \cdots, n\}$, and the pivots in the column-echelon form induce a bijective pairing $G \leftrightarrow H$, such that the following conditions hold:*

(i) *For every index i, the chains $\hat{\sigma}_1, \cdots, \hat{\sigma}_i$ form a basis of the chain group of K_i,*

(ii) *For every index $f \in F$, $\partial \hat{\sigma}_f$ is a cycle, i.e. $\partial \hat{\sigma}_f = 0$,*

(iii) *For every pair of indices (g, h) given by the pairing $G \leftrightarrow H$, $\partial \hat{\sigma}_h = \hat{\sigma}_g$ and so $\hat{\sigma}_g = 0$.*

Note that item (i) is equivalent to the assertion that the leading term of each simplicial chain $\hat{\sigma}_i$ is σ_i. Thus, the set F identifies the positive simplices which do not get paired, the set G identifies the positive simplices that do get paired, and the set H identifies the corresponding negative simplices.

The proof of the theorem is a sequence of simple lemmas. Although there is no record in the official literature, it can be found in some course notes such as [4, §12.5].

Example 11. Take the simplicial filtration $\mathcal{K}$ shown in Figure 1. From the reduced boundary matrix we can read off the interval decomposition of the persistent homology of $\mathcal{K}$:

$$\begin{aligned} \mathsf{H}_0(\mathcal{K}) &\cong \mathbb{I}[1, +\infty) \oplus \mathbb{I}[2, 4) \oplus \mathbb{I}[3, 5) \\ \mathsf{H}_1(\mathcal{K}) &\cong \mathbb{I}[6, 7) \end{aligned}$$

Observe that the essential feature $\mathbb{I}[1, +\infty)$ gives the homology of the solid triangle as expected.

The time complexity of the algorithm is at most cubic in the number m of simplices of K. To see this, observe that the j-th iteration of the `for` loop modifies only the j-th column of the matrix, therefore every column $k < j$ is considered at most once by the inner `while` loop at that iteration. Moreover, each column k can store the row number of its lowest entry once and for all after the k-th iteration of the `for` loop, so finding k such that $\mathsf{low}(k, R) = \mathsf{low}(j, R)$ at a subsequent iteration j can be done in time $O(j)$. Thus, the running time of the algorithm is indeed $O(m^3)$. A more careful analysis—see e.g. [9, §VII.2]—using a sparse matrix representation for M gives a tighter running-time bound[1] in $O(\sum_{j=1}^{m} p_j^2)$, where $p_j = j$ if simplex σ_j is unpaired and $p_j = j - i$ if σ_j is paired with σ_i. The quantity $\sum_{j=1}^{m} p_j^2$ is of the order of m^3 in the worst case, and a worst-case example on which the total running time is indeed cubic can be found in [12]. However, $\sum_{j=1}^{m} p_j^2$ is typically much smaller than that in practice, where worst-case scenarios are unlikely to occur and a near-linear running time is usually observed.

References

[1] Maurice Auslander. Representation theory of Artin algebras II. *Communications in Algebra*, 1:269–310, 1974.

[2] Gorô Azumaya. Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[3] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proc. Annual Symposium on Computational Geometry*, June 2014.

[4] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. Computational topology inference. Notes from the course *Computational Geometry Learning* given at the *Parisian Master of Research in Computer Science* (MPRI). `http://www-sop.inria.fr/geometrica/courses/supports/main.pdf`, October 2014.

[5] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Y. Oudot. The structure and stability of persistence modules. Research Report arXiv:1207.3674 [math.AT], July 2012.

[6] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, January 2007.

[7] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. Research Report arXiv:1210.0819 [math.RT], October 2012.

[8] V. de Silva, D. Morozov, and M. Vejdemo-Johansson. Dualities in persistent (co)homology. *Inverse Problems*, 27:124003, 2011.

[9] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction.* AMS Bookstore, 2010.

[10] Peter Gabriel. Unzerlegbare Darstellungen I. *Manuscripta Mathematica*, 6:71–103, 1972.

[11] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. Research Report arXiv:1106.5305 [cs.CG], June 2011.

---

[1] This bound assumes the dimension of each simplex to be constant.

[12] Dmitriy Morozov. *Homological Illusions of Persistence and Stability.* Ph.D. dissertation, Department of Computer Science, Duke University, 2008.

[13] James R Munkres. *Elements of algebraic topology*, volume 2. Addison-Wesley Reading, 1984.

[14] Cary Webb. Decomposition of graded modules. *Proceedings of the American Mathematical Society*, 94(4):565–571, 1985.