PERSISTENT HOMOLOGY, MATROIDS AND COBORDISMS

İSMAIL GÜZEL AND ATABEY KAYGUN

Abstract. The homological information about a filtered simplicial complex over the poset of positive real numbers is often presented by a barcode which depicts the evolution of the associated Betti numbers. However, there is a wonderfully complex combinatorics associated with the homology classes of a filtered complex, and one can do more than just counting them over the index poset. Here, we show that this combinatorial information can be encoded by filtered matroids, or even better, by rooted forests. We also show that these rooted forests can be realized as cobordisms.

Introduction

Persistent homology is a tool data scientists recently started using to understand collections of data points embedded in a metric space [6, 11]. The basic process is one that from a given set of data points one constructs a complex filtered over the poset of positive real numbers (the poset of scale parameters), and then calculates the Betti numbers of this complex at different scale parameters. Then one expects that the gradual evolution of the Betti numbers of the filtered complex does yield some information about the topology of the subspace from which our data points are sampled.

The fact that one can capture the homological and homotopical invariants of a space from a finite collection of points sampled from the space under certain guarantees is an old idea [3]. However, in the absence of any information whether these guarantees are satisfied, one has to construct a sample of invariants from available local information by playing with the notion of proximity via a scale parameter that we alluded above. Since we do not know which range of scale parameters truly capture the topological invariants of the underlying space, one must calculate these invariants at different scale parameters and investigate how these different calculations fit with each other.

It is perhaps a historical coincidence that the development of persistent homology mirrors that of the ordinary homology. In the beginning topologists calculated homology as Betti numbers, and it was Emmy Noether who first observed that these homological invariants had to be considered as abelian groups [15]. Similarly, in the beginning the practitioners represented persistent homology as barcodes, which are records of how Betti numbers evolve as the scale parameter varies. Then it is clear that we must consider persistent homology as a filtered abelian group should we make the same leap.

In this paper, we propose a new combinatorial description, which we call the cophenetic matroid, for homological groups that vary over a scale parameter. This description squarely fits between barcodes and filtered vector spaces. We also show that one can represent these filtered matroids via rooted forests that come from specific cobordisms of disjoint unions of spheres.

1 We review some of these complexes in Section 1.8.

2 Even though there is now a plethora of different representations, such as persistence diagrams [9], landscapes [4], images [1], terraces [18], entropy [17] and curves [8]; they all are derived from the barcode representation [6, 11].

3 Filtered matroids have been used before by Henselman and Ghrist in [14], however, their aim was to develop and implement efficient algorithms for computing cosheaf homology. Moreover, they still used barcodes as their medium of representation for their persistent homology calculations.

1
Plan of the article. In Section 1 we recall basic facts about matroids, filtered simplicial complexes, and persistent homology. In Section 2 we define filtered matroids and show that every matroid filtered over the set of positive numbers can be represented by a rooted forest. In Section 3 we define the cophenetic matroid that we are going to use to represent persistent homology. We then recover the non-archimedean metric we defined in [13] using the cophenetic matroid. Finally in Section 4 we construct the cobordism that lies underneath the cophenetic matroid of a filtered complex.

1. Preliminaries

1.1. Posets and order ideals. A poset is a set $P$ together with an anti-symmetric reflexive and transitive relation $\leq_P$. A function $f : P \rightarrow Q$ between two posets is called order preserving if $x \leq_P y$ implies $f(x) \leq_Q f(y)$ for every $x, y \in P$. Given two order preserving maps $f, g : P \rightarrow Q$ we say that $g$ dominates $f$ if $f(x) \leq_Q g(x)$ for every $x \in P$. A subset $\mathcal{I}$ of a poset $P$ is called an order ideal if for every $y \in \mathcal{I}$ and $x \in P$ if $x \leq y$ then $x \in \mathcal{I}$.

1.2. Matroids. A matroid $M$ is a pair $(E, \mathcal{I})$ where $E$ is a non-empty set and $\mathcal{I}$ is a non-empty order ideal in the poset $(2^E, \subseteq)$ such that for every $A, B \in \mathcal{I}$ with $|A| < |B|$ there is an element $x \in B \setminus A$ such that $\{x\} \cup A \in \mathcal{I}$. Elements of $\mathcal{I}$ are called independent sets.

1.3. The rank function of a matroid. Let $M$ be a matroid on a finite ground set $E$. The rank $r(X)$ of a subset $X \subseteq E$ is the cardinality of the largest independent set contained in $X$. In other words

$$r(X) = \max\{|A| \in \mathbb{N} \mid A \subseteq X \text{ and } A \in \mathcal{I}\}$$

Notice that the rank function $r : 2^E \rightarrow \mathbb{N}$ is order preserving and is dominated by the cardinality function $|\cdot| : 2^E \rightarrow \mathbb{N}$.

We can convert matroids to rank functions and vice versa. To show this we need the following definition: A poset map $r : 2^E \rightarrow \mathbb{N}$ is called semimodular or submodular if

$$r(A \cup B) \leq r(A) + r(B) - r(A \cap B)$$

for every $A, B \in 2^E$. A submodular map $r$ is called modular if the inequality is replaced with an equality, i.e. when $r$ satisfies the inclusion/exclusion principle.

Theorem 1.1 ([12] Chap. 2.5, pg 69). Let $E$ be a finite set and let $r : 2^E \rightarrow \mathbb{N}$ be a poset map dominated by the cardinality function. Then $r$ is the rank function of a matroid if and only if $r$ is submodular.

Thus Theorem 1.1 gives us license to replace any matroid with its rank function, and vice versa.

1.4. Morphisms of matroids. Assume $(E, r_E)$ and $(F, r_F)$ are two matroids given by rank functions. A set map $f : E \rightarrow F$ is called a morphism of matroids if $r_F(f(A)) \leq r_E(A)$ for every finite subset $A$ of $E$. One can easily see that identity map is a morphism of matroids, and composition of any two morphism is again a morphism. So, we have a category of matroids.

1.5. Induced matroids. Let $(E, r_E)$ be a matroid and assume $\pi : F \rightarrow E$ is any function. Let us define

$$\pi^* r_E(A) := r_E(\pi(A))$$

for every finite subset $A$ of $F$. The following Lemma is pretty straightforward and its proof is left to the reader.

Lemma 1.2. The pair $(F, r_F)$ is a matroid and $\pi : (E, r_E) \rightarrow (F, \pi^* r_E)$ is a morphism of matroids.
1.6. **Simplicial complexes.** Given a space $X$, a simplicial complex $\mathcal{K}$ in $X$ is an order ideal in $(2^X, \subseteq)$. If $\mathcal{K}$ consists of finite sets we write

$$\mathcal{K}_n = \{ x \in \mathcal{K} \mid |x| = n \}$$

for every $n \in \mathbb{N}$.

1.6.1. **Clique complex.** Let $K$ be a graph. The clique complex of $K$ is a simplicial complex $\mathcal{L}$ such that if any set of vertices $\{x_0, \ldots, x_n\}$ forms a clique, i.e. when all possible edges between these vertices are in $K$, then the simplex $[x_0, \ldots, x_n]$ is in $\mathcal{L}$.

1.6.2. **The nerve of a topological space.** Let $X$ be a topological space and let $\mathcal{U} = \{U_i \subseteq X : i \in I\}$ be a covering $X$ by open sets indexed by a set $I$. The nerve $N(\mathcal{U})$ of the covering $\mathcal{U}$ is defined as the simplicial complex $C(\mathcal{U})$ whose vertices (i.e. 0-th skeleton) is $\mathcal{U}$ as a set. For every $k \geq 1$, the ordered collection $[U_{i_0}, \ldots, U_{i_k}]$ is going to be $k$-simplex if the intersection $\bigcap_{j=0}^k U_{i_j}$ is non-empty.

The nerve of a topological space $X$, rather a suitable covering $\mathcal{U}$, is a useful construction since one can recover the homotopy/homology type of $X$ from $N(\mathcal{U})$.

**Proposition 1.3** ([?]). Let $X$ be a topological space with an open cover $\mathcal{U} = \{U_i : i \in I\}$. Assume that the intersection of elements of any subset of $\mathcal{U}$ is empty or contractible. Then, the space $X$ and its nerve $N(\mathcal{U})$ are homotopy equivalent.

1.7. **Filtered complexes.** Let $(P, \leq)$ be an indexing poset. We call a collection $(\mathcal{K}_\varepsilon)_{\varepsilon \in P}$ of simplicial sets indexed by $P$ as a filtered simplicial complex over $P$ if for every comparable pair of element $\varepsilon \leq \eta$ in $P$ we have a morphism of simplicial sets of the form $t_{\varepsilon, \eta} : \mathcal{K}_\varepsilon \rightarrow \mathcal{K}_\eta$ such that

$$t_{\eta, \nu} \circ t_{\varepsilon, \eta} = t_{\varepsilon, \nu}$$

for every $\varepsilon \leq \eta \leq \nu$ in $P$. One can also define a filtered complex as a functor $\mathcal{K}$ from the poset $P$ to the category of simplicial complexes.

1.8. **A zoo of filtered complexes.** Now, let $(X, d)$ be a metric space and let $D \subseteq X$ be a point cloud in $X$. Let us use $B_\varepsilon(x)$ for the open ball of radius $\varepsilon$ centred at $x \in D$.

1.8.1. **Čech complex.** The Čech complex $C_\varepsilon(D)$ associated with $D$ at a scale parameter $\varepsilon$ is defined to be the nerve $N(\mathcal{U}_\varepsilon)$ of the covering

$$\mathcal{U}_\varepsilon = \{B_\varepsilon(x) \mid x \in D\}$$

1.8.2. **Vietoris-Rips complex.** The Vietoris-Rips complex $R_\varepsilon(D)$ of $D$ is defined to be the simplicial complex whose vertices are all points in $D$ that are at most $\varepsilon$ apart. In other words

$$R_\varepsilon(D) = \{ \sigma \subset D \mid d(x, y) \leq \varepsilon, \text{ for all } x, y \in \sigma \}$$

The clique complex of a graph $K$ is an example of Vietoris-Rips complex if we consider a graph as a metric space via the geodesic distance and set $\varepsilon = 1$.

1.8.3. **Delaunay complex.** The Voronoi region $R(x)$ of a point $x \in D$ is defined as the points in $X$ that are closest to $x$. Formally,

$$R(x) = \{ y \in X \mid x \in \text{argmin}_{z \in D} d(z, y) \}$$

The Delaunay complex is the nerve of the covering $\mathcal{R} = \{R(x) \mid x \in D\}$ of $D$ given by Voronoi regions.
1.8.4. **Alpha complex.** Let $\varepsilon > 0$. The restricted Voronoi region of a point $x$ is the intersection of the Voronoi region $R(x)$ and the open ball $B_\varepsilon(x)$. The alpha complex $A_\varepsilon$ is the nerve of the covering given by the restricted Voronoi regions

$$A_\varepsilon = \{ B_\varepsilon(x) \cap R(x) \mid x \in D \}$$

The alpha complex grows with $\varepsilon$. For instance, $A_0 = \emptyset$ and if $\varepsilon$ is big enough $A_\varepsilon$ coincides with the Delaunay complex. Moreover, unlike the Rips and the Čech complexes, the dimension of the alpha complex is restricted to the dimension of the space the points are embedded in given that the points are in general position. For example, the dimension of the alpha complex of a set of points in $\mathbb{R}^2$ cannot exceed 2 whenever none three points are collinear.

1.9. **The persistent homology.** Assume $\mathcal{K}$ is a simplicial complex. Let $C_n(\mathcal{K})$ be the $k$-span of the simplices in $\mathcal{K}$ for every $n \geq 0$ where we define $d_n = \sum_{i=0}^n (-1)^i \partial_i$. Because of the simplicial identities we have $d_{n-1}d_n = 0$ and we define

$$Z_n = \ker(d_n), \quad B_n = \text{im}(d_{n+1}), \quad H_n(\mathcal{K}) = Z_n / B_n$$

for every $n \geq 0$. The vector space $Z_n$ is the space of cycles, $B_n$ is the space of boundaries, and $H_n(\mathcal{K})$ is the homology of $\mathcal{K}$.

In persistent homology, the simplicial complexes we have are filtered over the poset $\mathbb{R}_+$ with its natural order. Then for a filtered complex $(\mathcal{K}_\varepsilon)_{\varepsilon \in P}$, the $k$-th persistent homology of the filtered complex is defined as

$$\text{PH}_k(\mathcal{K}):= \{ H_k(\mathcal{K}_\varepsilon) \}_{\varepsilon \in P}$$

together with the collection of $k$-linear maps of the form $\psi^k_{\varepsilon, \eta}: H_k(\mathcal{K}_\varepsilon) \to H_k(\mathcal{K}_\eta)$ induced by the structure maps of the filtration $\iota_{\varepsilon, \eta}: \mathcal{K}_\varepsilon \to \mathcal{K}_\eta$ for all $k \in \mathbb{N}$ and $\varepsilon \leq \eta$ in $\mathbb{R}_+$.

1.10. **Bar codes of persistent homology.** Since persistent homology works with filtered complexes over $\mathbb{R}_+$, for each cycle $\gamma \in Z^k_{\mathcal{K}}$ there is an interval that records the life-time of $\gamma$, i.e. the interval on which $\gamma$ is non-trivial as $\varepsilon$ ranges from 0 to $\infty$. We say $\gamma$ is born at $\varepsilon = b$ when the homology class $[\gamma] \in H_k(\mathcal{K}_b)$ is not in the image of $\psi^k_{\varepsilon, b}$ for every $\varepsilon < b$. Similarly, we say $\gamma$ dies at $\varepsilon = d$ if $\psi^k_{b, \varepsilon}([\gamma]) = 0$ for every $\varepsilon > d$.

![Figure 1. An example barcode.](image)

To illustrate the life-times of cycles, we use barcodes as introduced by Carlsson et.al. [6] and Ghrist [11]. In a barcode, we place the basis vectors for the homology on the vertical axis whereas the horizontal axis represents the life span of each basis element in terms of the scale parameter.
\( \varepsilon \). When we draw the vertical line at a particular \( \varepsilon_i \), the number of intersecting line segments in a barcode is the dimension of the corresponding homology group, i.e. the Betti number, for that parameter \( \varepsilon_i \). See Figure 1.

2. Filtered Matroids and Rooted Forests

2.1. Irreducible sets in a matroid. Assume \((E, r_E)\) is a matroid. We call a subset \( A \subseteq E \) as irreducible if \( r_E(A) = |A| - 1 \) and for every proper subset \( B \) of \( A \) we have \( r_E(B) = |B| \).

**Proposition 2.1.** Given any finite subset \( X \subseteq E \) with \( r_E(X) < |X| \), there are irreducible subsets \( A_1, \ldots, A_n \) such that \( X = \bigcup_{i=1}^{n} A_i \).

**Proof.** We give the proof on induction on the size of \( X \). For \( |X| = 1 \), \( X \) is already an irreducible set and the statement is obviously true. So, let us assume the statement holds for every \( k \leq n \) and let \( |X| = n + 1 \) with \( r_E(X) \leq n \). Take any element \( x \in X \) and consider the set \( \mathcal{U} \) of all subsets \( A \subseteq X \) such that \( x \in A \) and \( A \) is irreducible. Since \( \mathcal{U} \) is a non-empty finite poset, there are maximal elements. Let \( Y \in \mathcal{U} \) be such a maximal set. If it is already \( Y = X \) one can stop. Otherwise, we remove \( x \) from \( X \) and proceed by induction. \( \square \)

2.2. Filtered Matroids. Let \( P \) be an indexing poset. A filtered matroid \((M_\varepsilon)_{\varepsilon \in P}\) is a set of pairs \((E_\varepsilon, r_\varepsilon) : 2^{E_\varepsilon} \rightarrow \mathbb{N}\) indexed by \( P \) where \( E_\varepsilon \) is a set and \( r_\varepsilon \) is a rank function. We must also have functions \( \psi_{\eta,\varepsilon} : E_\varepsilon \rightarrow E_\eta \) that satisfy the conditions

\[
\psi_{\eta,\varepsilon} \circ \psi_{\eta,\eta} = \psi_{\eta,\varepsilon} \quad r_\eta(\psi_{\eta,\eta}(A)) \leq r_\varepsilon(A)
\]

for every finite set \( A \subseteq E_\varepsilon \) and for every \( \varepsilon \leq \eta \leq \nu \) in \( P \).

Here is another interpretation: Let us view \( P \) as a category such that there is a unique morphism \( x \to y \) whenever \( x \leq y \) in \( P \). Then a filtered matroid is a functor from \( P \) into the category of matroids.

2.3. Ramification of irreducible sets. Let us assume \( P \) is an indexing poset and let \((E_\varepsilon, r_\varepsilon)\) be a filtered matroid over \( P \) with structure maps \( \psi_{\eta,\varepsilon} : E_\varepsilon \rightarrow E_\eta \) for every \( \varepsilon \leq \eta \) in \( P \). A irreducible set \( A \subseteq E_\varepsilon \) is said to be ramified at \( \eta > \varepsilon \) if \( r_\eta(\psi_{\eta,\varepsilon}(A)) < r_\varepsilon(A) \).

**Theorem 2.2.** Assume \((E_\varepsilon, r_\varepsilon, \psi_{\eta,\varepsilon})\) is a filtered matroid over the poset \( \mathbb{R}_+ \). For every \( \varepsilon \) and for every irreducible subset \( A \) of \( E_\varepsilon \), one can write the ramification information as a finite rooted tree whose edges are labeled by irreducible sets.\( \square \)

**Proof.** Assume \( A \) ramified at \( \eta > \varepsilon \), i.e. \( r_\eta(\psi_{\eta,\varepsilon}(A)) \leq r_\varepsilon(A) = |A| - 1 \). Then by Proposition 2.1 we can write \( \psi_{\eta,\varepsilon}(A) \) as a union of irreducible sets. Since \( A \) is finite, \( A \) can only ramify finitely many times. \( \square \)

The rooted tree of an irreducible set \( A \) is going to be called the ramification tree or the ramification dendrogram of the irreducible set \( A \).

2.4. An example. For every \( \varepsilon \in [0, \infty) \) let us define \( s_\varepsilon : \mathbb{R}^n \to \mathbb{R}^n \) as

\[
s_\varepsilon(x_1, \ldots, x_n) = \begin{cases} (x_1, \ldots, x_n) & \text{if } 0 \leq \varepsilon < 1 \\ (0, \ldots, 0, x_i, \ldots, x_n) & \text{if } i \leq \varepsilon < i + 1 \\ (0, \ldots, 0) & \text{if } \varepsilon \geq n + 1 \end{cases}
\]

Let \( F_n \) be the set of all finite subsets of \( \mathbb{R}^n \) and define \( r_\varepsilon : F_n \to \mathbb{N} \) by

\[
r_\varepsilon(A) = \dim s_\varepsilon(A)
\]

(2.1)
One can check that this is a filtered matroid. Consider
\[ A = \{(1, 1, 1), (1, 1, 2, 2), (1, 2, 3, 3), (3, 5, 6, 6)\} \]
where we have \( r_0(A) = 3 \) and every subset of \( A \) has rank 3 which means \( A \) is irreducible. But
\[ s_1(A) = \{(0, 1, 1, 1), (0, 1, 2, 2), (0, 2, 3, 3), (0, 5, 6, 6)\} \]
has rank 2, and therefore, is not irreducible. We can write \( s_1(A) \) as a union of irreducible sets of maximal rank 2
\[ s_1(A) = s_1(A_1) \cup s_1(A_2) \]
where
\[ A_1 = \{(1, 1, 1, 1), (1, 1, 2, 2), (1, 2, 3, 3)\} \quad A_2 = \{(1, 1, 2, 2), (1, 2, 3, 3), (1, 5, 6, 6)\} \]
These irreducible sets further reduce at \( \varepsilon = 2 \) and we split
\[ s_2(A_1) = s_2(A_{11}) \cup s_2(A_{12}) \quad s_2(A_2) = s_2(A_{12}) \cup s_2(A_{22}) \]
where
\[ A_{11} = \{(1, 1, 1, 1), (1, 1, 2, 2)\} \quad A_{12} = \{(1, 1, 2, 2), (1, 2, 3, 3)\} \quad A_{22} = \{(1, 2, 3, 3), (3, 5, 6, 6)\} \]
and each set has rank 1. These sets preserve their ranks until \( \varepsilon = 4 \) and after \( \varepsilon \geq 4 \) all subset reduce to 0. Thus we can write the tree as shown in Figure 2.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.png}
\caption{The rooted tree representation of the matroid given in Subsection 2.4.}
\end{figure}

3. Persistent Homology and Filtered Matroids

3.1. Multi-dimensional persistence and the no-go theorem of Bauer et.al. Let \( n \geq 1 \) and let us consider the poset
\[ \mathbb{R}^n_\varepsilon = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n\} \]
together with the partial simplicial ordering \((x_1, \ldots, x_n) \leq (y_1, \ldots, y_n)\) if \( x_i \leq y_i \) for every \( 1 \leq i \leq n \). Given a filtered simplicial complex \( \mathcal{K}_\varepsilon \) over \( \mathbb{R}^n_\varepsilon \), one may try to construct barcodes similar to the ordinary bar codes of [6,11]. Barcodes are complete invariants due to the fact that representation theory of the poset \( \mathbb{R}^n_+ \) is rather simple. However, no such simple representations exist for filtered complexes over \( \mathbb{R}^n_\varepsilon \) since the representation theory of the poset \( \mathbb{R}^n_+ \) and its discretization \( \mathbb{N}^n \) are both wild for \( n \geq 2 \) by [2].

3.2. Carlsson-Zomorodian rank function. The no-go result of [2] forces us to come up with new representations to depict evolutions of persistent homology classes over a scale parameter. One such example is by Carlsson and Zomorodian [5].

Assume \( M_\varepsilon \) is a \( \mathbb{R}^n_\varepsilon \)-filtered vector space where we assume \( \dim_\varepsilon(M_\varepsilon) \) is finite for every \( \varepsilon \in \mathbb{R}^n_\varepsilon \). In other words, we have finite dimensional vector spaces \( M_\varepsilon \) for each \( \varepsilon \in \mathbb{R}^n_+ \) together with structure maps \( \psi_{\varepsilon, \eta} : M_\varepsilon \to M_\eta \) for every \( \varepsilon \leq \eta \). Then the Carlsson-Zomorodian rank function of \( M \) is defined to be
\[ \rho(\varepsilon, \eta) := \dim_\varepsilon \psi_{\varepsilon, \eta}(M_\varepsilon) \]
for every \( \varepsilon \leq \eta \in \mathbb{R}^n_+ \) [5, Definition 6].
3.3. **Carlsson-Zomorodian matroid.** There is a finer invariant than Carlsson-Zomorodian rank function given by a filtered matroid.

**Proposition 3.1.** Given any filtered finite dimensional vector spaces \( (M_\varepsilon) \), the function \( r_\varepsilon \) defined as

\[
(3.1) \quad r_\varepsilon(A) = \dim_{\mathbb{R}} \text{Span}_{\mathbb{R}}(A)
\]

for every finite subset \( A \) of \( M_\varepsilon \) yields a filtered matroid.

**Proof.** The function \( r_\varepsilon \) is dominated by the cardinality function, and it satisfies

\[
\dim_{\mathbb{R}} \psi_{\varepsilon,\eta}(\text{Span}_{\mathbb{R}}(A)) \leq \dim_{\mathbb{R}} \text{Span}_{\mathbb{R}}(A)
\]

for every \( \varepsilon \leq \eta \), and thus, the collection \( (r_\varepsilon)_{\varepsilon \in \mathbb{R}^+} \) is a filtered matroid. \( \square \)

3.4. **Cophenetic matroid.** From this point onward, we work with the poset \( \mathbb{R}_+ \), and a filtered simplicial complex \( (\mathcal{K}_\varepsilon)_{\varepsilon \in \mathbb{R}_+} \) such that structure morphisms are inclusions. Recall that for this filtered complex we have cycles \( Z^k_\varepsilon := \ker(d^k_\varepsilon) \) and boundaries \( B^k_\varepsilon := \text{im}(d^k_{k+1}) \) and homology groups \( H_k(\mathcal{K}_\varepsilon) := Z^k_\varepsilon / B^k_\varepsilon \) for every \( k \in \mathbb{N} \) and \( \varepsilon \in \mathbb{R}_+ \). We also have connecting linear maps \( \psi_{\varepsilon,\eta}^k : Z^k_\varepsilon \rightarrow Z^k_\eta \) and \( \psi_{\varepsilon,\eta}^k : B^k_\varepsilon \rightarrow B^k_\eta \) for every pair \( \varepsilon \leq \eta \) and for every \( k \in \mathbb{N} \). Note that since \( \mathcal{K}_\varepsilon \subseteq \mathcal{K}_\eta \) for every \( \varepsilon \leq \eta \), the induced maps \( \psi_{\varepsilon,\eta}^k \) on cycles and boundaries are also monomorphisms. However, even if this is the case, the induced maps in homology need not be monomorphisms.

Let us write \( F^c_k \) for the set of all finite subsets of \( Z^c_k \). For every \( A \in F^c_k \) define

\[
(3.2) \quad c^k(A) = \dim(\text{Span}_{\mathbb{R}}(A) + B^c_k) - \dim B^c_k
\]

\[
(3.3) \quad = \dim(\text{Span}_{\mathbb{R}}(A)) - \dim(\text{Span}_{\mathbb{R}}(A) \cap B^c_k)
\]

Notice that \( c^k_\varepsilon \) is a poset map and is dominated by the cardinality function and we have

\[
(3.4) \quad c^k_\eta(\psi_{\varepsilon,\eta}(A)) \leq c^k_\varepsilon(A)
\]

for every \( \eta \geq \varepsilon \) and \( A \in F^c_k \). The function \( c^k_\varepsilon \) is called the *cophenetic rank function* of the filtered complex \( \mathcal{K}_\varepsilon \).

**Theorem 3.2.** The cophenetic rank function \( c^k_\varepsilon : F^c_k \rightarrow \mathbb{N} \) is submodular for every \( \varepsilon \in \mathbb{R}_+ \) and for every \( k \in \mathbb{N} \). Thus by Theorem 1.1 for every \( k \geq 0 \) there is a filtered matroid \( (M^k_\varepsilon)_{\varepsilon \in \mathbb{R}_+} \) of the filtered simplicial complex \( (\mathcal{K}_\varepsilon)_{\varepsilon \in \mathbb{R}_+} \).

**Proof.** Given a finite set \( A \) in \( Z^c_k \) its cophenetic rank \( c^k_\varepsilon(A) \) is the dimension of \( \text{Span}_{\mathbb{R}}(A) \) in the quotient vector space \( H_k(\mathcal{K}_\varepsilon) = Z^c_k / B^c_k \). Now, apply Lemma 1.2 \( \square \)

The matroid \( (M^k_\varepsilon)_{\varepsilon \in \mathbb{R}_+} \) given in Theorem 3.2 is called the *k-th cophenetic matroid* of a filtered simplicial complex \( (\mathcal{K}_\varepsilon)_{\varepsilon \in \mathbb{R}_+} \).

3.5. **An example.** Consider the configuration of points given in Figure 3. Assume we put a filtration where at \( \varepsilon = 0 \) we have disjoint points, and at \( \varepsilon = 1 \) the smaller triangles \( DEF, GHI \) and \( JKL \) are formed. Then at \( \varepsilon = 2 \) the large triangle \( ABC \) is formed, and at \( \varepsilon = 3 \) we fill-in the region between the large triangle \( ABC \) and the three smaller triangles. Finally at \( \varepsilon = 4, 5, 6 \) we fill-in the smaller triangles \( DEF, GHI \) and \( JKL \) in order.

Consider the set of first homology classes \( X = \{ABC, DEF, GHI, JKL\} \) that forms as an independent set at \( \varepsilon = 2 \). But at \( \varepsilon = 3 \) when we fill-in the region between \( ABC \) and the smaller
triangles they become linearly dependent, and we get an irreducible set. As we kill the smaller triangles we get

\[ X = \{ABC, GIH, JKL\} \cup \{DEF\} \]
\[ = \{ABC, JKL\} \cup \{GIH\} \cup \{DEF\} \]
\[ = \{ABC\} \cup \{JKL\} \cup \{GIH\} \cup \{DEF\} \]

as unions of irreducible sets. We represent these splittings as a tree in Figure 4.

**Figure 4.** Tree representation of the cophenetic matroid of Example 3.5

### 3.6. Cophenetic distance

Now, for each pair of cycles \( \alpha \) and \( \beta \) in \( Z^k_E \) representing classes in \( H_k(\mathcal{X}_\varepsilon) \), one can test the rank of the pair \( \{\alpha, \beta\} \) at every \( \eta > \varepsilon \). If the rank of the pair \( \{\psi^k_{\varepsilon,\eta}(\alpha), \psi^k_{\varepsilon,\eta}(\beta)\} \) is less than 2, then we will say that the cycles \( \alpha \) and \( \beta \) merged at time \( \eta \). Thus we can define \( k \)-th homological cophenetic distance

\[ d_k(\alpha, \beta) = \inf \{\eta - \varepsilon > 0 \mid c^k_{\eta}(\{\psi^k_{\varepsilon,\eta}(\alpha), \psi^k_{\varepsilon,\eta}(\beta)\}) < 2\} \]

for every \( \alpha, \beta \in H_k(\mathcal{X}_\varepsilon) \) and for every \( k \geq 0 \).

**Proposition 3.3** ([13]). The cophenetic distance \( d_k \) on \( H_k(\mathcal{X}_\varepsilon) \) is a non-archimedean metric for every \( \varepsilon \geq 0 \) and for every \( k \geq 0 \).

**Proof.** Assume \( \alpha, \beta, \gamma \in Z_k(\mathcal{X}_\varepsilon) \). Assume, by way of contradiction, that

\[ d_k(\alpha, \beta) > \max(d_k(\alpha, \gamma), d_k(\gamma, \beta)) \]
This means there are indices $\eta > \mu$ such that the pair $(\alpha, \beta)$ becomes linearly dependent in $H_k(\mathcal{X}_\eta)$ while the pairs $(\alpha, \gamma)$ and $(\gamma, \beta)$ become linearly dependent at an earlier time in $H_k(\mathcal{X}_\mu)$. Then there are non-zero scalars $a, b \in k$ such that
\[
\alpha = a \gamma, \quad \beta = b \gamma \quad \text{which implies} \quad b \alpha = a \beta
\]
in $H_k(\mathcal{X}_\mu)$ which is a contradiction since $\alpha$ and $\beta$ become linearly dependent at a later time $\eta > \mu$.

\[\Box\]

4. Cobordisms

4.1. Hurewicz map. Assume that our data set $D$ is sampled from a manifold $M$ embedded in $\mathbb{R}^n$. Assume also we created a filtered simplicial complex $(\mathcal{X}_\epsilon)$ from $D$. Since we work with filtered complexes $(\mathcal{X}_\epsilon)_{\epsilon \in \mathbb{R}_+}$, the corresponding vector spaces of cycles $(Z^\epsilon_k)_{\epsilon \in \mathbb{R}_+}$ and boundaries $(B^\epsilon_k)_{\epsilon \in \mathbb{R}_+}$ are also filtered.

First, we recall the following version of the rational Hurewicz Theorem:

**Proposition 4.1** ([10] [16]). Assume $M$ is a simply connected topological space with $\pi_n(M) = 0$ for $1 \leq n < r$. Then the rational Hurewicz map $\pi_n(M) \otimes \mathbb{Q} \to H_n(M) \otimes \mathbb{Q}$ is an isomorphism for $1 \leq n < 2r - 1$ and is a surjection for $n = 2r - 1$.

We use Proposition 4.1 as follows.

**Proposition 4.2.** Let us assume $\pi_n(C) = 0$ for each connected component $C$ of $M$ for all $0 \leq n < r$ for some $r$. Then for all $0 \leq n \leq 2r - 1$ all cycles in $Z^\epsilon_n$, in particular, every boundary in $B^\epsilon_n$ comes from an embedded $n$-sphere in $M$.

**Proof.** If a connected component $C$ is simply connected, i.e. when $r = 1$, then we use Proposition 4.1. If $C$ fails to be simply-connected then the classical Hurewicz map $\pi_1(C) \to H_1(C)$ is already surjective for every path connected component $C$ of $M$.

\[\Box\]

4.2. Dendrograms of irreducible sets as cobordisms of spheres. Assume our data $D$ is sampled from a manifold $M$ that satisfies the hypothesis of Proposition 4.2. Assume also that we constructed a filtered complex $(\mathcal{X}_\epsilon)$ out of $D$.

**Theorem 4.3.** The ramification tree of every irreducible set $A$ in $H_n(\mathcal{X}_\epsilon)$ can be represented by a $n + 1$ dimensional cobordism of disjoint $n$-spheres for every $0 \leq n \leq 2r - 1$.

**Proof.** For every $0 \leq n < 2r - 1$, and irreducible collection of homology $n$-cycles $A$ there is a $n + 1$-sphere with $k$-punctures such that punctures represent classes in $A$ and the $n + 1$-sphere implements the linear dependence of elements in $A$. This is because every cycle $\alpha \in Z^\epsilon_n$ and boundary $\beta \in B^\epsilon_{n+1}$, and their every scalar multiple, is represented with a sphere via the Hurewicz map. If a collection $A$ in $H_n(\mathcal{X}_\epsilon)$ is irreducible, the elements in $A$ represented by $n$-spheres have to be linearly dependent given by a boundary which is a $n + 1$-sphere. The result follows.

\[\Box\]

**References**

[1] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, The Journal of Machine Learning Research, 18 (2017), pp. 218–252.

[2] U. Bauer, M. B. Botnan, S. Oppermann, and J. Steen, Cotorsion torsion triples and the representation theory of filtered hierarchical clustering, Advances in Mathematics, 369 (2020), p. 107171.

[3] K. Borsuk, On the imbedding of systems of compacta in simplicial complexes, Fundamenta Mathematicae, 35 (1948), pp. 217–234.
[4] P. Bubenik, *Statistical topological data analysis using persistence landscapes*, The Journal of Machine Learning Research, 16 (2015), pp. 77–102.

[5] G. Carlsson and A. Zomorodian, *The theory of multidimensional persistence*, Discrete Comput. Geom., 42 (2009), pp. 71–93.

[6] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, *Persistence barcodes for shapes*, International Journal of Shape Modeling, 11 (2005), pp. 149–187.

[7] F. Chazal and B. Michel, *An introduction to topological data analysis: Fundamental and practical aspects for data scientists*, Frontiers in Artificial Intelligence, 4 (2021).

[8] Y.-M. Chung and A. Lawson, *Persistence curves: A canonical framework for summarizing persistence diagrams*, arXiv preprint arXiv:1904.07768, (2019).

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, *Stability of persistence diagrams*, Discrete & computational geometry, 37 (2007), pp. 103–120.

[10] M. Dyer, *Rational homology and Whitehead products*, Pacific J. Math., 40 (1972), pp. 59–71.

[11] R. Ghrist, *Barcodes: the persistent topology of data*, Bulletin of the American Mathematical Society, 45 (2008), pp. 61–75.

[12] G. Gordon and J. McNulty, *Matroids: a geometric introduction*, Cambridge University Press, 2012.

[13] I. Güzel and A. Kaygun, *A new non-archimedean metric on persistent homology*, Computational Statistics, 37 (2022), pp. 1963–1983.

[14] G. Henselman and R. Ghrist, *Matroid filtrations and computational persistent homology*, 2016.

[15] P. Hilton, *A brief, subjective history of homology and homotopy theory in this century*, Mathematics Magazine, 61 (1988), pp. 282–291.

[16] S. Klaus and M. Kreck, *A quick proof of the rational Hurewicz theorem and a computation of the rational homotopy groups of spheres*, Math. Proc. Cambridge Philos. Soc., 136 (2004), pp. 617–623.

[17] E. Merelli, M. Rucco, P. Sloot, and L. Tesei, *Topological characterization of complex systems: Using persistent entropy*, Entropy, 17 (2015), pp. 6872–6892.

[18] C. Moon, N. Giansiracusa, and N. A. Lazar, *Persistence terrace for topological inference of point cloud data*, Journal of Computational and Graphical Statistics, 27 (2018), pp. 576–586.

Email address: iguzel@itu.edu.tr

Email address: kaygun@itu.edu.tr

Department of Mathematics, Istanbul Technical University, Istanbul, Turkey.