Persistent homology: definition and computation

Žiga Virk

November 26, 2021

The concept of persistent homology along with its variations is at the forefront of topological data analysis. Mathematically speaking, persistent homology is an obvious extension of homology: the functoriality of homology is applied to a sequence of inclusions. The resulting structure is, somewhat surprisingly, not harder to compute than ordinary homology. When coupled with the standard constructions of complexes, persistent homology contains not only topological but also geometric information.

We will start this chapter by explaining geometric intuition on persistent homology. We will continue by presenting formal and convenient visualisation techniques. We will conclude with a fairly simple algorithm for computation one could call a "labelled matrix reduction".

1 Definition

We start by describing a geometric intuition of persistent homology. Given a "growing" simplicial complex persistent homology describes the evolution of its holes. As an illustrative example we consider four simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ of Figures 1 and 2.
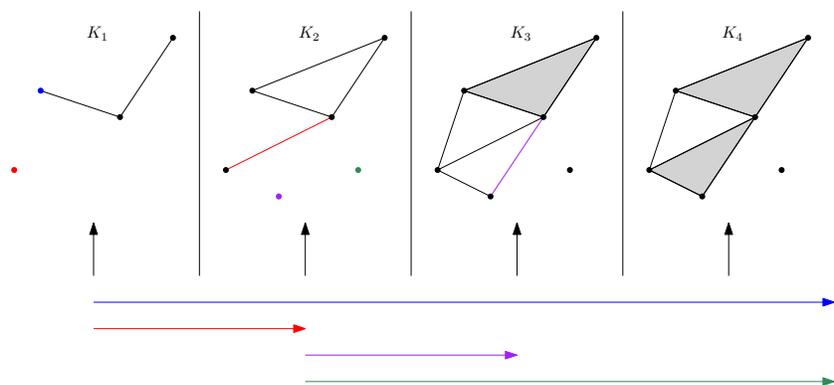

Figure 1: Nested simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ are divided by vertical lines. The horizontal arrows below are called "**bars**" and form a **barcode**. They indicate the persistence of zero-dimensional homology classes: components. The left endpoint of each bar corresponds to the birth complex of a component. The right endpoint of each bar corresponds to the terminal complex of a component. The color of each bar also appears on one vertex (the representative of the component) and potentially on one edge (the edge, that kills the component).

Here is how we interpret the corresponding zero-dimensional barcode[1] described by Figure 1:

K_1 : There are two components of K_1. This fact is visualised by the fact that there are two bars (blue and red) starting at that time. The corresponding homology generators[2] (points) are colored accordingly.

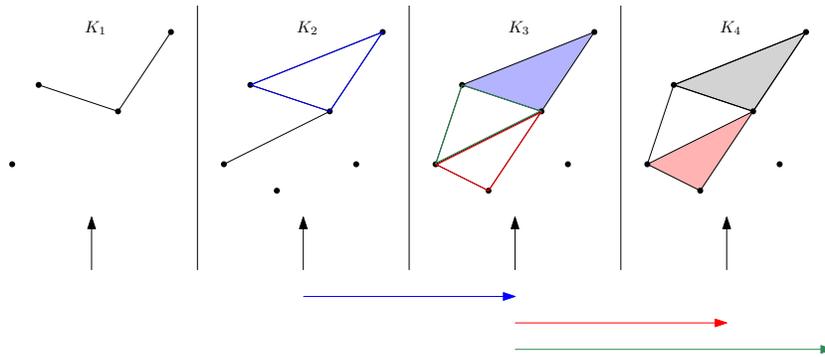K_2 : There are three components of K_1: this fact is visualised by the

[1] I.e., the indicated evolution of the components.

[2] Note that the generator of the red component is unique. On the other hand, we could have chosen any vertex of the other component as a generator and color it blue.

fact that there are three bars (blue, purple and green) passing from that time on. The corresponding homology generators of the new components are colored accordingly. However, the two components of K_1 merge, which we interpret as one of the components of K_1 disappearing. We declare[3] that the component disappearing is the red component, which is visualised by the fact that the red bar terminates just before K_2. The edge making the connection between the two components is colored in red.

K_3 : The purple component terminates by connecting to the blue component via two edges, one of which is indicated by the purple color.

K_4 : There is no change in components as compared to K_3, both bars are passing through to infinity.

[3] As the components appeared at the same time, we might as well have chosen to have the blue bar terminated and keep the red bar going. The uncolored barcode would have remained the same. However, whenever there is a merger of components with different birth times we act according to the **elder rule**: the older component survives. This will be apparent at K_3. The reader may rest assured this is not a product of discrimination but rather a rule that is consistent with the mathematical structure of persistence (especially the interleaving and stability) that will be described later.

Figure 2: Nested simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ and the corresponding one-dimensional homology barcode.

In a similar fashion we interpret the corresponding one-dimensional barcode[4] described by Figure 2:

[4] I.e., the indicated evolution of the holes.

K_1 : There are no holes and hence no bars passing on.

K_2 : A blue hole appears inducing a blue bar.

K_3 : The blue hole becomes trivial by the blue triangle and hence the blue bar terminates. However, two new holes appear, the red one and the green one. Consequently, there are two bars passing from K_3 on.

K_4 : The red hole becomes trivial while the green hole lives on forever, just as the corresponding bar.

The goal of this chapter is to present the theoretical background formalizing the presented geometric idea of persistent homology, and to introduce the computational procedure to obtain the barcodes.

Formal definition

We first formally introduce a filtration: a nested sequence of ever larger simplicial complexes modelling a growing simplicial complex.

Definition 1.1. *Let K be a simplicial complex. A (discrete) **filtration** of K is a sequence of subcomplexes*

$$K_1 \leq K_2 \leq \ldots \leq K_m = K.$$

An example of a filtration is given in Figures 1 and 2.

Persistent homology measures how homology elements[5] persist[6] through steps of a filtration. A filtration

$$K_1 \leq K_2 \leq \ldots \leq K_m = K$$

of a simplicial complex K can be expressed as a sequence of natural inclusion maps denoted by map[7] $i_{\_,\_}$:

$$K_1 \overset{i_{1,2}}{\hookrightarrow} K_2 \overset{i_{2,3}}{\hookrightarrow} \ldots \overset{i_{m-1,m}}{\hookrightarrow} K_m = K.$$

Given a field $\mathbb{F}$ and $q \in \{0,1,2,\ldots\}$ we can apply homology $H_q(\_;\mathbb{F})$ to obtain a sequence[8] of homology groups connected by maps:

$$H_q(K_1;\mathbb{F}) \xrightarrow{(i_{1,2})_*} H_q(K_2;\mathbb{F}) \xrightarrow{(i_{2,3})_*} \ldots \xrightarrow{(i_{m-1,m})_*} H_q(K_m;\mathbb{F}) = H_q(K;\mathbb{F})$$

Definition 1.2. *Assume K is a simplicial complex, $\mathbb{F}$ is a field, and $q \in \{0,1,2,\ldots\}$. Given a filtration*

$$K_1 \leq K_2 \leq \ldots \leq K_m = K$$

*of K, the corresponding q-dimensional **persistent homology** groups with coefficients in $\mathbb{F}$ are images of the maps*

$$(i_{s,t})_* \colon H_q(K_s;\mathbb{F}) \to H_q(K_t;\mathbb{F})$$

*for all $0 \leq s \leq t \leq m$. The corresponding ranks $\beta^q_{s,t} = \mathrm{rank}(i_{s,t})_*$ are called **persistent Betti numbers**.*

As is the case with the ordinary homology, each persistent homology group is determined[9] up to isomorphism by its Betti number. A single filtration results in a table of persistent Betti numbers.

Example 1.3. *Given any field $\mathbb{F}$ the following are the tables of the zero-dimensional and one-dimensional persistent Betti numbers of the filtration of Figure 3:*

[5] I.e., components, holes, etc.

[6] I.e., remain non-trivial

[7] For example, $i_{s,t}\colon K_s \hookrightarrow K_t$.

[8] By the functoriality of the homology we have $(i_{u,t})_* \circ (i_{s,u})_* = (i_{s,t})_*$.

☞ In each step of a filtration we add simplices. The addition of a single d-dimensional simplex in one step may either "kill" a non-trivial homological element of dimension $d-1$ or create a non-trivial homological element of dimension d.

☞ Note that $\beta^q_{s,t}$ is a non-increasing function in t and a non-decreasing function in s.

[9] While the rank of $(i_{s,t})_*$ determines the image of the map $i_{s,t}$ up to isomorphism, it does not determine a specific $\beta_{s,t}$-dimensional subspace of $H_q(K_t;\mathbb{F})$. In this aspect persistent homology as a specific subgroup of $H_q(K_t;\mathbb{F})$ contains more information that persistent Betti numbers, i.e., its basis consists of homology representatives spanning the persistent homology group.

$$\beta^0_{s,t} \rightarrow$$

$s \backslash t$	1	2	3	4
1	2	1	1	1
2	/	3	2	2
3	/	/	2	2
4	/	/	/	2

$$\beta^1_{s,t} \rightarrow$$

$s \backslash t$	1	2	3	4
1	0	0	0	0
2	/	1	0	0
3	/	/	2	1
4	/	/	/	1

Table 1: The table of persistent Betti numbers corresponding to the filtration of Figure 3. The diagonal entries coincide with the Betti numbers of the corresponding stages of the filtration. The sub-diagonal entries are undefined.

Let us demonstrate how to interpret these numbers geometrically:

- $\beta^0_{2,3} = 2$ *means that two of the different components of K_2 are still disconnected from each other in K_3.*

- $\beta^1_{3,4} = 1$ *roughly means that only one homologically non-trivial loop of K_3 is still[10] homologically non-trivial in K_4.*

- $\beta^1_{2,3} = 0$ *means all one-dimensional homology elements in $H_1(K_2; \mathbb{F})$ are homologically trivial in K_3.*

[10] A mathematically correct statement would be: the space of one-dimensional homology elements in $H_1(K_4; \mathbb{F})$ which have represenatives in $C_1(K_3; \mathbb{F})$ is of dimension one.
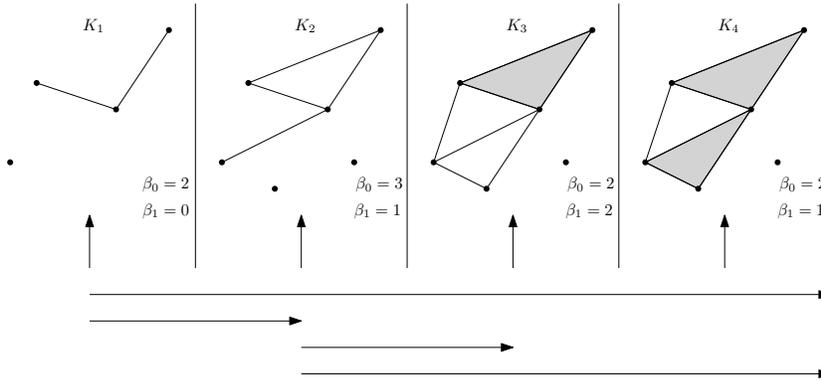

Figure 3: A filtration $K_1 \leq K_2 \leq K_3 \leq K_4$ along with the corresponding Betti numbers of each of the stage and the zero-dimensional barcode.

While the tables of persistent Betti numbers are useful, there are other ways to visualize the evolution of homology groups through a filtration. One such visualization we have already presented is the barcode.

2 Visualization

Throughout this section we fix a field $\mathbb{F}$, $q \in \{0, 1, \ldots\}$, a filtration

$$K_1 \leq K_2 \leq \ldots \leq K_m = K,$$

and $1 \leq s < t \leq m$.

Barcodes

The barcodes have been geometrically introduced above. In this subsection we will provide their formal definition.

Persistent Betti number $\beta_{s,t}^q$ represents the dimension of the subspace of homology elements in K_t that have a representative in K_s. Putting it differently, $\beta_{s,t}^q$ indicates the dimension of the collection[11] of homology elements in K_s that are still non-trivial in K_t in the sense that $\beta_{s,t}^q = \dim H_q(K_s; \mathbb{F})/\ker(i_{s,t})_*$. Barcodes as indicated above however have a more specific information: a bar $[s, t)$ represents a homology element that is born precisely at s and terminates precisely at t. Let us phrase this formally:

1. The number of bars containing s and passing through t equals[12] $\beta_{s,t}^q$.

2. Homology **born** at s is defined as[13] $H_q(K_s; \mathbb{F})/(\operatorname{Im} i_{s-1,s})_*$. Its dimension is $\beta_{s,s} - \beta_{s-1,s}$ and represents the number of bars starting at s.

3. Homology **terminating** at t is defined as $\ker(i_{t-1,t})_*$. Its dimension[14] is $\beta_{t-1,t-1} - \beta_{t-1,t}$ and represents the number of bars terminating at t.

4. Quantity $\beta_{s,t} - \beta_{s-1,t}$ represents[15] the dimension of homology born at s which is still alive at t. It represents the number of bars starting at s which are passing through t.

5. Quantity $\mathbf{n_{s,t}} = \beta_{s,t-1} - \beta_{s-1,t-1} - (\beta_{s,t} - \beta_{s-1,t})$ represents[16] the dimension of homology born at s which terminates at t. It represents the number of bars starting at s and terminating at t.

6. We additionally define $\mathbf{n_{s,\infty}} = \beta_{s,m} - \beta_{s-1,m}$, which represents the dimension of homology born at s which is still alive at the end of filtration.

The q-dimensional **barcode**[17] consists of intervals[18] of the form

i. $[s, t)$ for $1 \le s < t \le m$, and

ii. $[s, \infty)$ for $1 \le s < m$.

A barcode can have multiple[19] copies[20] of each interval. Fixing $1 \le s < t \le m$:

- The number of the intervals $[s, t)$, denoted by $n_{s,t}$.

- The number of the intervals $[s, \infty)$, denoted by $n_{s,\infty}$.

Example 2.1. *We again turn our attention the familiar filtration in Figure 3. From the table on the right we can deduce that $n_{2,3} = 1 - 1 - (2 - 3) = 1$ and as a result there is 1 bar of the form $[2, 3)$, as displayed in the figure. In a similar fashion we compute $n_{1,2} = n_{1,\infty} = n_{2,\infty} = 1$ and $n_{1,3} = n_{1,4} = n_{2,4} = n_{3,4} = n_{3,\infty} = n_{4,\infty} = 0$.*

[11] This collection is formally not a linear subspace. In a formal setting we represent it as the quotient linear subspace appearing at the end of the sentence.

[12] Through the rest of the section we will drop the superscript q indicating the fixed dimension.

[13] For formal reasons we define $(i_{0,t})_*$ to be the trivial map.

[14] Using the fact that $\ker(i_{t-1,t})_* \cong H_q(K_{t-1}; \mathbb{F})/\operatorname{Im}(i_{t-1,t})_*$.

[15] Compare to the interpretation of persistent Betti numbers above. Also note that $\beta_{s,t} - \beta_{s-1,t} = \dim((\operatorname{Im} i_{s,t})_*/\operatorname{Im}(i_{s-1,t})_*)$, i.e., the dimension of the homology elements in $H_q(K_t; \mathbb{F})$ that have a representative in K_s module the ones that have a representative in K_{s-1}.

[16] Observation 4. interprets this formula as [the dimension of homology born at s which is alive at $t - 1$] - [the dimension of homology born at s which is still alive at t] .

[17] ...of the chosen filtration with coefficients in $\mathbb{F}$...

[18] In the setting of a barcode these intervals will be called **bars**.

[19] ...or none...

[20] Alternatively, we could think of the barcode as the collection of all possible intervals of the forms (i) and (ii), each with an assigned multiplicity from $\{0, 1, 2, \ldots\}$.

$\beta_{s,t}^0 \to$	$s \;\backslash\; t$	1	2	3	4
	1	2	1_-	1_+	1
	2	/	3_+	2_-	2
	3	/	/	2	2
	4	/	/	/	2

Computing $n_{2,3}$ of Figure 3: the coloring corresponds to the defining formula and the subscripts to the signs within.

A barcode represents the persistences of homology elements. The longer a bar, the longer the corresponding homology element persists. In most settings the longer persistence of a homology element also means higher importance[21]. However, there are also settings in which information is contained in shorter bars, especially when the bars are numerous and specifically distributed.

[21] I.e., a more prominent topological feature.

Persistence diagrams

Another established method of visualisation of persistent homology is persistence diagram defined as follows. Given a barcode as defined above we can think of an interval $[s, t)$ as a pair of numbers and visualize[22] it as a point $(s, t) \in \mathbb{R}^2$. A point of the form (s, ∞) obviously can't be drawn in a plane so we choose a y-coordinate above k, perhaps most conveniently as $k + 1$, to act as a representative of ∞, i.e., a bar $[s, \infty)$ corresponds to a point $(s, k + 1)$. Each point (s, t) of a persistence diagram has an assigned **multiplicity** $n_{s,t}$, which represents the number of bars of the form $[s, t)$. In the case of $(s, k + 1)$, the multiplicity is $n_{s,\infty}$.

The result is a collection of weighted points in the plane called **persistence diagram**. An example is provided in Figure 4.

[22] Just as there can be more bars with the same endpoints in a persistence diagram, there can be more copies of the same point visualized at the same location in a persistence diagram. While multiple such intervals can be visualized in a vertical stack, the same can not be done with points. For this reason we always consider a point (s, t) in a persistence diagram as a weighted point with weight (multiplicity) $n_{s,t}$.
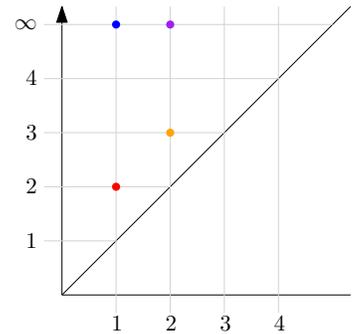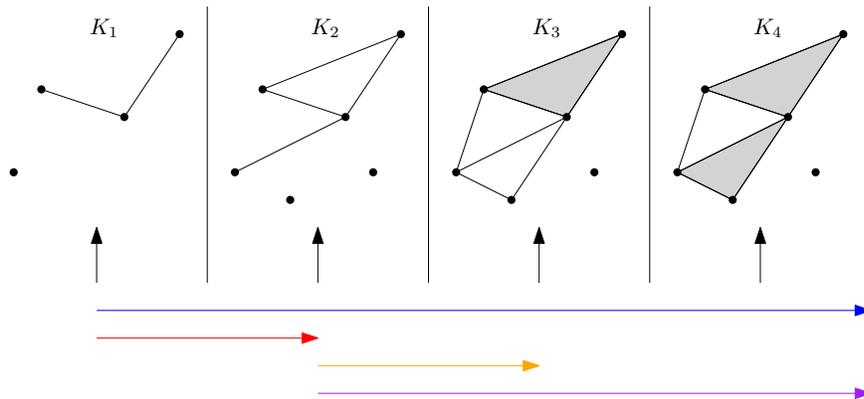

Figure 4: A filtration along with the corresponding zero-dimensional barcode and persistence diagram. The colors of bars match the colors of the corresponding points in the persistence diagram.

☞ Theoretically speaking, if there existed bars $[s, s)$ of length zero, then these would have been the shortest bars. They would have corresponded to diagonal points (s, s). This point of view will come handy in the next chapter in the context of stability.

A barcode encodes precisely the same information as a persistence diagram. While the persistence of a bar is measured by its length, the persistence of a point on a persistence diagram is measured by its distance from the **diagonal** $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$. All points of a persistence diagram lie above Δ.

Persistence diagrams are often the method choice of visualization when it comes to representations of persistent homology. Especially when the number of points and bars is large, their distribution seems to be well represented by persistence diagrams. On the other hand, when the number of points and bars is low, a barcode is often more

descriptive.

The fundamental lemma of persistent homology

Numbers $n_{s,t}$ are defined using persistent Betti numbers $\beta_{s,t}$. It turns out that the reverse expression also exists.

Lemma 2.2. *[The fundamental lemma of persistent homology]*

$$\beta_{s,t} = \sum_{s' \leq s,\ t' > t} n_{s,t}$$

The formula in the lemma can be verified explicitly. However, the statement is apparent from the definitions, as

- $\beta_{s,t}$ represents the homology born at s or before and terminating after t;

- $n_{s,t}$ represents the homology born precisely at s and terminating precisely at t.

Lemma 2.2 implies that the information encoded in a barcode or in a persistence diagram is precisely the same as the information encoded by persistent Betti numbers.

3 Computation

While the multiplicities $n_{s,t}$ of points of persistence diagrams are formally expressed by persistent Betti numbers, there is an algorithm to obtain them directly without referring to the Betti numbers and the corresponding $k(k+1)/2$ ranks of maps. In this section we will present perhaps the simplest[23] version of the algorithm, which is also the most illustrative. We will proceed in two steps:

- compute the matrix reduction, and

- extract the persistent homology.

We will conclude the section with an example.

Throughout this section we fix a field $\mathbb{F}$ and filtration

$$K_1 \leq K_2 \leq \ldots \leq K_m = K.$$

Parameter $q \in \{0, 1, \ldots\}$ will denote the dimension of a considered object.

Matrix reduction

This part could be called an annotated matrix reduction using only column operation from the left.

⚠ In this setting, condition $t' > t$ implies that t' as an index in n could also take the value ∞.

☞ Lemma 2.2 has a geometric interpretation in the context of persistence diagrams, see Figure 5. It essentially states that $\beta_{s,t}$ is the sum of all multiplicities of points of a persistence diagram, which lie in the upper-left quadrant $[0, s] \times (t, \infty]$ with the apex at (s, t). In the context of this interpretation, the formula for multiplicity $n_{s,t} = \beta_{s,t-1} - \beta_{s-1,t-1} - \beta_{s,t} + \beta_{s-1,t}$ is the expression of the square $(s-1, s] \times [t-1, t)$ in terms of such quadrants.
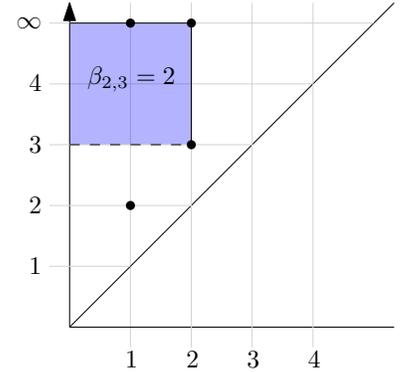

Figure 5: The sum of multiplicities of points in the blue quadrant with apex $(2, 3)$ is $\beta_{2,3}$ by Lemma 2.2. It equals 2 as the point $(2, 3)$ is not contained in it, see also the table in Example 1.3.

[23] There exist many improvements of this algorithm which may significantly improve the computing time.

1. **Order simplices consistently with the filtration**. For each q order all q-simplices in an order in which they appear in the filtration. If more simplices appear at the same time their internal order is immaterial[24]

2. For each q **construct the boundary matrix M_q using the order** chosen in 1 to label columns and rows.

3. For each q **reduce M_q from the left** using a **single type of column operations**: the addition of an $\mathbb{F}$-multiple of any of the previous[25] columns to a treated column. Specifically, starting with the leftmost column go through all the columns by passing to the right and for each column:

   (a) Determine the pivot[26].

   (b) If any of the previous columns on the left has a pivot in the same row, subtract the appropriate multiple of that column so that the pivot of the current column either disappears or its location is moved up.

   (c) Repeat as long as there are matching pivots on the left.

   For each q the resulting matrix is denoted by M_q'. Each of its columns is either trivial or has a pivot, whose row is unique amongst all pivots.

Extracting persistence

At this point we have sufficient information to extract homology of K_m from the number of pivots[27]. However, we can also use the locations of pivots to extract numbers $n_{s,k}$ required to construct the barcode and persistence diagram. In order to explain the extraction process we first recall the incremental expansion we have discussed in a previous chapter.

Given a simplicial complex, an addition of a single q-simplex can change the homology in two ways:

- If its boundary is a linear combination of boundaries of other q-simplices[28], then the simplex gives birth to a non-trivial q-dimensional homology element. In this case we call the simplex a **birth simplex**.

- If its boundary is *not* a linear combination of boundaries of other q-simplices[29], then the simplex kills to a non-trivial $(q-1)$-dimensional homology element. In this case we call the simplex a **terminal simplex**.

[24] It would eventually effect the obtained critical simplices and representatives, but not the persistent homology.

[25] In the chosen order from 1.

[26] The lowest non-trivial entry in the column

[27] Note that the rank of a matrix is the number of its pivots in a reduced form, and the ranks themselves suffice to compute the Betti numbers.

[28] I.e., if, after adding the simplex to the boundary matrix as the rightmost column, its column gets reduced to the trivial column by the above reduction.

[29] I.e., if, after adding the simplex to the boundary matrix as the rightmost column, its column does not reduce to the trivial column.

A filtration can be considered to be a sequence[30] of incremental expansions. At each stage of the filtration we may assume we first add all vertices according to the ordering in 1., the all edges, etc. Combining such an ordering through all stages we get a sequence of incremental expansions inducing boundary matrices M_q and their reduced forms M_q'.

Based on such an ordering each simplex of K is either a terminal simplex or a birth simplex. We are now in a position to **extract persistent Betti numbers**:

- For each terminal q-simplex τ there exists a paired birth $(q-1)$-simplex σ, which is the label of the pivot in the column τ. Such a pair **induces a bar** $[s, t)$ in the corresponding barcode or, equivalently, **a point** (s, t) in the corresponding persistence diagram, where s, t are the stages of the filtration at which σ and τ appear[31].

- Each birth simplex which is not paired to a terminal simplex **induces a bar** $[s, \infty)$ in the corresponding barcode or, equivalently, **a point** $(s, m+1)$ in the corresponding persistence diagram, where s is the stage of the filtration at which σ appears.

As a result we obtain a barcode and a persistence diagram as demonstrated in the example in the last subsection.

Representatives

Occasionally we are also interested in homology representatives of the bars and points of persistence diagrams. These can be extracted from the reduction process. In this subsection we present the most direct way of generating representatives. Given a bar with the birth simplex σ and the terminal simplex τ we define:

- The **birth representative** of σ as the chain formulated[32] by the reduction of the column corresponding to σ to the zero column in the column reduction scheme. In particular, if the linear combination turning column σ into the zero column in our column reduction scheme is encoded in terms of columns as $\partial\sigma - \sum_i \lambda_i \partial\sigma_i = 0$, then the birth representative is $\alpha = \sigma - \sum_i \lambda_i \sigma_i$. The birth representative gives a homology $[\alpha]$ class that is born[33] by the addition of σ.

- The **terminal representative** is encoded by the column corresponding[34] to τ in the reduced matrix.

The birth representative and the terminal representative typically do not represent the same homology class. The birth representative may net even be a good representative of the corresponding bar in the sense that it may remain homologically non-trivial beyond[35] the

appearance of the corresponding terminal simplex. On the other hand infinite intervals do not have a terminal representative. As a result we define the **representative of a bar** as follows:

1. If the bar is a finite interval, the representative of the bar is the terminal representative.

2. If the bar is an infinite interval, the representative of the bar is the birth representative.

This choice of representatives is algebraically sound in the sense that the representatives form a basis of the elementary intervals of the decomposition described in the structure theorem for persistent homology, a result we discuss in details in the next chapter. This statement includes the fact that the lifespan of each representative matches the lifespan of the corresponding bar, and that the representatives are linearly independent[36] at all times.

In practice we sometimes deviate from the algebraically orthodox choice by making an exception when declaring the representatives of 0-dimensional bars: we choose the birth representative as a bar representative even if the bar is bounded. Let us explain this geometrically motivated declaration on the example of the next subsection, where pair $(\langle b\rangle, \langle a, b\rangle)$ induces a 0-dimensional bar. Sometimes we would geometrically like to think of this bar as a representation of the component containing b merging with a larger component, hence the choice of the birth representative $\langle b\rangle$ which fits into this geometric intuition. However, we should be aware that homological element $[\langle b\rangle]$ does not become trivial[37] after adding $\langle a, b\rangle$. In terms of homology the appearance of $\langle a, b\rangle$ identifies[38] $[\langle a\rangle] = [\langle b\rangle]$ rather than sets $[\langle b\rangle] = 0$.

Example

As an example we compute persistent homology of our standard example, see Figure 4. The annotation of simplices we will be using is provided in Figure 6. The chosen order is apparent from the following boundary matrices, in which vertical and horizontal lines divide simplices from different stages of filtration.

✍ There is no guarantee that these representatives are geometrically the most convenient. There are more involved ways of obtaining representatives that optimize given criterion function. For example, we may want to obtain the shortest 1-dimensional representatives, etc..

☞ Let us prove that the lifespan of the homology class of the terminal representative β corresponds to the lifespan of the corresponding bar:

- $[\beta]$ appears by the time σ appears by construction;

- if a representative β' of $[\beta]$ appeared before σ, then the column corresponding to τ could have been reduced further to β' thus eliminating the pivot labeled as σ, a contradiction.

- $[\beta]$ becomes trivial by the time τ emerges by definition;

- if $[\beta]$ became trivial sooner, its expression as a boundary could be used to reduce the τ column to the zero column, a contradiction.

[36] ...or trivial beyond their lifespans

[37] The terminal representative $\langle b\rangle - \langle a\rangle$ does become trivial. In fact $[\langle b\rangle]$ never becomes trivial.

[38] In this sense the terminal representative tells us which two components merge.
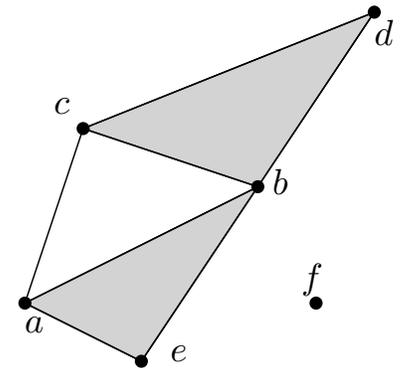
Figure 6: The annotation of simplices of K.

$$
M_1 = \begin{array}{c} \\ \langle a\rangle \\ \langle b\rangle \\ \langle c\rangle \\ \langle d\rangle \\ \langle e\rangle \\ \langle f\rangle \end{array}
\begin{array}{c} \langle b,c\rangle \quad \langle b,d\rangle \quad \langle a,b\rangle \quad \langle c,d\rangle \quad \langle a,c\rangle \quad \langle a,e\rangle \quad \langle b,e\rangle \\
\left(\begin{array}{ccc|ccc|cc}
 & & -1 & & -1 & -1 & \\
-1 & -1 & 1 & & & & -1 \\
1 & & & -1 & 1 & & \\
 & 1 & & 1 & & & \\
\hline
 & & & & & 1 & 1 \\
 & & & & & &
\end{array}\right)
\end{array}
$$

$$M_2 = M_2' = \begin{array}{c} \\ \langle b,c\rangle \\ \langle b,d\rangle \\ \langle a,b\rangle \\ \langle c,d\rangle \\ \langle a,c\rangle \\ \langle a,e\rangle \\ \langle b,e\rangle \end{array} \begin{array}{c} \langle b,c,d\rangle \quad \langle a,b,e\rangle \\ \left(\begin{array}{c|c} 1 & \\ -1 & \\ & 1 \\ \boxed{1} & \\ & \\ & -1 \\ & \boxed{1} \end{array} \right) \end{array}$$

✎ Green entries are the pivots.

We now perform the labelled matrix reduction as described above.

☞ In the matrices below a blue column is modified using red columns.

$$M_1 = \begin{array}{c} \\ \langle a\rangle \\ \langle b\rangle \\ \langle c\rangle \\ \langle d\rangle \\ \langle e\rangle \\ \langle f\rangle \end{array} \begin{array}{c} \langle b,c\rangle \ \langle b,d\rangle \ \langle a,b\rangle \ \langle c,d\rangle \ \langle a,c\rangle \ \langle a,e\rangle \ \langle b,e\rangle \\ \left(\begin{array}{ccc|cc|ccc} & & -1 & & -1 & -1 & \\ -1 & -1 & 1 & & & & -1 \\ 1 & & & -1 & 1 & & \\ & 1 & & 1 & & & \\ \hline & & & & & 1 & 1 \\ & & & & & & \end{array} \right) \end{array}$$

✎ $\partial\langle c,d\rangle = \partial\langle b,d\rangle - \partial\langle b,c\rangle$.

$$\begin{array}{c} \\ \langle a\rangle \\ \langle b\rangle \\ \langle c\rangle \\ \langle d\rangle \\ \langle e\rangle \\ \langle f\rangle \end{array} \begin{array}{c} \langle b,c\rangle \ \langle b,d\rangle \ \langle a,b\rangle \ \langle c,d\rangle \ \langle a,c\rangle \ \langle a,e\rangle \ \langle b,e\rangle \\ \left(\begin{array}{cc|cc|ccc} & & -1 & & -1 & -1 & \\ -1 & -1 & 1 & & & & -1 \\ 1 & & & & 1 & & \\ & 1 & & & & & \\ \hline & & & & & 1 & 1 \\ & & & & & & \end{array} \right) \end{array}$$

✎ $\partial\langle a,c\rangle = \partial\langle a,b\rangle + \partial\langle b,c\rangle$.

$$\begin{array}{c} \\ \langle a\rangle \\ \langle b\rangle \\ \langle c\rangle \\ \langle d\rangle \\ \langle e\rangle \\ \langle f\rangle \end{array} \begin{array}{c} \langle b,c\rangle \ \langle b,d\rangle \ \langle a,b\rangle \ \langle c,d\rangle \ \langle a,c\rangle \ \langle a,e\rangle \ \langle b,e\rangle \\ \left(\begin{array}{cc|cc|ccc} & & -1 & & & -1 & \\ -1 & -1 & 1 & & & & -1 \\ 1 & & & & & & \\ & 1 & & & & & \\ \hline & & & & & 1 & 1 \\ & & & & & & \end{array} \right) \end{array}$$

✎ $\partial\langle b,e\rangle = \partial\langle a,e\rangle - \partial\langle a,b\rangle$.

$$M_1' = \begin{array}{c} \\ \langle a\rangle \\ \langle b\rangle \\ \langle c\rangle \\ \langle d\rangle \\ \langle e\rangle \\ \langle f\rangle \end{array} \begin{array}{c} \langle b,c\rangle \ \langle b,d\rangle \ \langle a,b\rangle \ \langle c,d\rangle \ \langle a,c\rangle \ \langle a,e\rangle \ \langle b,e\rangle \\ \left(\begin{array}{cc|cc|ccc} & & -1 & & & -1 & \\ -1 & -1 & \boxed{1} & & & & \\ \boxed{1} & & & & & & \\ & \boxed{1} & & & & & \\ \hline & & & & & \boxed{1} & \\ & & & & & & \end{array} \right) \end{array}$$

We can now extract the barcode from the birth-terminal pairs and unpaired birth simplices. We start by extracting the zero-dimensional barcode from the pivots of M_1' and unpaired vertices.

- Pairs $(\langle c \rangle, \langle b,c \rangle)$ and $(\langle d \rangle, \langle b,d \rangle)$ provide no contribution[39].

- Pair $(\langle b \rangle, \langle a,b \rangle)$ induces a 0-dimensional (component) bar[40] $[1,2)$ representedby $[\langle b \rangle]$.

- Pair $(\langle e \rangle, \langle a,e \rangle)$ induces a 0-dimensional (component) bar $[2,3)$ represented by $[\langle e \rangle]$.

- Vertices a and f are unpaired and thus induce 0-dimensional bars $[1,\infty)$ (generated by $[\langle a \rangle]$) and $[2,\infty)$ (generated by $[\langle f \rangle]$).

We next extract the one-dimensional barcode from the pivots of M_2' and unpaired edges.

- Pair $(\langle c,d \rangle, \langle b,c,d \rangle)$ induces a 1-dimensional bar $[2,3)$ represented[41] by $[\langle c,d \rangle - \langle b,d \rangle + \langle b,c \rangle]$.

- Pair $(\langle b,e \rangle, \langle a,b,e \rangle)$ induces a 1-dimensional bar $[3,4)$ represented by $[\langle b,e \rangle - \langle a,e \rangle + \langle a,b \rangle]$.

- Edge $\langle a,c \rangle$ is unpaired and thus induce the 1-dimensional bar $[3,\infty)$ generated by $[\langle a,c \rangle - \langle a,b \rangle - \langle b,c \rangle]$.

Computational tricks

We conclude by mentioning a trick that speeds up the computation of persistent homology. It based on an observation that boundary matrices M_i that are being reduced in the reduction process are not completely independent of each other.

If the reduction process of M_q reduces the column corresponding to q-simplex τ to a non-trivial column, we can extract the following information

1. τ is a terminal simplex and hence the row corresponding to τ in M_{q+1} will have been reduced to the zero-row, which means we can set it to zero immediately.

2. The pivot location reveals the corresponding birth simplex σ. As a result the column corresponding to σ in M_{q-1} will have been reduced to the zero-column and can hence be set to zero immediately.

Hence a single reduction of a column in M_q corresponding to a terminal simplex also reveals a zero-row in M_{q+1} and a zero-column in M_{q-1}. Of course, this information can't be of much help if it has already been extracted from previous reductions. For this reasons the matrices M_q can be reduced in the order of decreasing dimension: this way no column in M_{q-1} has been reduced by the time M_q has been reduced and as a result we avoid reducing almost half[42] the columns resulting in a significant speedup.

[39] Formally, they contribute the empty interval $[1,1)$ as all involved simplices appear at K_1.

[40] Recall that $\langle b \rangle$ appears at K_1 while $\langle a,b \rangle$ appears at K_2, hence the values of the endpoints.

✎ In this example the 1-dimensional birth and terminal representatives coincide. This is not generally the case.

[41] Recall that $\langle c,d \rangle$ appears at K_2 while $\langle b,c,d \rangle$ appears at K_3, hence the values of the endpoints. The linear combination that made the column corresponding to $\langle c,d \rangle$ trivial in M_1' was $\partial \langle c,d \rangle - \partial \langle b,d \rangle + \partial \langle b,c \rangle$ and hence the representative.

☞ Overview: reducing a τ-column to a non-zero column reveals:

- τ is a terminal simplex;
- τ-row is trivial;
- pivot label σ is a birth simplex;
- σ-column is trivial.

[42] This estimate depends on a filtration but seems to hold for most of the practical cases.

4 Concluding remarks

Recap (highlights) of this chapter

- Persistent homology;

- Barcode;

- Persistence diagram;

- Computing persistent homology.

Background and applications

Persistent homology is perhaps the most popular and fruitful construction of topological data analysis. For the past two decades it has been inspiration to extensive theoretical and practical treatments, spanning from purely mathematical theoretical foundations[43] to computable aspects and applications in numerous fields of science and engineering. When coupled with standard constructions of complexes, persistent homology contains information about geometry of data. As such, the method is applied whenever the geometric shape of data is thought to contain significant information.

Applications include de-noising schemes, dimension reduction schemes, feature extraction methods, and specific data analysis of materials, molecular structures, medical images, weather patterns, etc.

The combinatorial treatment of this chapter will be complemented by further properties in the following chapter. While the definition of persistent homology could have been expressed using coefficients in an Abelian group, the existence of the visualizations[44] and efficient implementations[45] crucially depend on the structure of a field.

Appendix: Zig-Zag persistence and Multi-parameter persistence

In this appendix we will sketch the ideas of two generalizations of the standard persistent homology as presented throughout the chapter. In both cases the generalization refers to the type of filtration used.

The first generalization is based on the Zig-Zag filtration. While a standard filtration models a growing simplicial complex, a Zig-Zag filtration models a changing simplicial complex, in which the simplices may be appearing or disappearing.

> **Definition 4.1.** *Let K be a simplicial complex. A **Zig-Zag filtration** of K is a sequence of subcomplexes $K_1, K_2, \ldots, K_m$ of K, such that for each $i \in \{1, 2, \ldots, m-1\}$ either $K_i \leq K_{i+1}$ or $K_{i+1} \leq K_i$.*

[43] We will mention two ideas of generalizations of the standard persistent homology in the appendix.

[44] I.e., barcodes and persistence diagrams.
[45] I.e., matrix reductions.

☞ In case $L_1, L_2, \ldots, L_m$ are subcomplexes of a simplicial complex L not satisfying the condition of a Zig-Zag filtration, and we still want to compute a meaningful Zig-Zag homology, a standard way to construct a corresponding Zig-Zag filtration is to connect them either by unions or intersections:

$$K_1 \hookrightarrow K_1 \cup K_2 \hookleftarrow K_2 \hookrightarrow \ldots \hookleftarrow K_m,$$

$$K_1 \hookleftarrow K_1 \cap K_2 \hookrightarrow K_2 \hookleftarrow \ldots \hookrightarrow K_m.$$

Interestingly enough, while the two options induce generally different barcodes, they encode precisely the same information.

For example, a Zig-Zag filtration may be of the following sort:

$$K_1 \hookrightarrow K_2 \hookleftarrow K_3 \hookleftarrow K_4 \hookrightarrow K_5 \hookleftarrow K_6 \hookrightarrow K_7.$$

It turns out that even in this setting there exists an algorithm based on matrix reductions which will produce a well defined barcode[46] describing what is called a **Zig-Zag persistence**. An example is displayed in Figure 7.
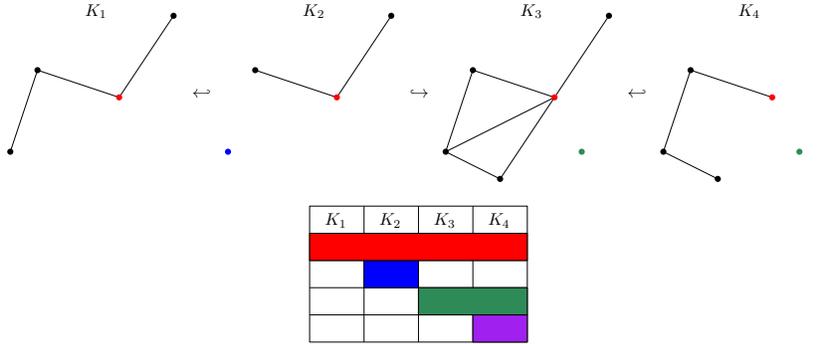
Figure 7: A Zig-Zag filtration and the corresponding zero-dimensional barcode, visualized as a table for technical reasons (i.e., the absence of a designated direction of all arrows as endpoints of bars). In the same way we could have presented the barcodes of the ordinary persistent homology as well.

The second generalization is based on the multi-parameter filtration. While a standard persistent homology models a one-parameter[47] growth of a simplicial complex, a multi-parameter filtration models growth with more degrees of freedom. For our demonstrative purposes it suffices to formally introduce only a 2-parameter filtration.

[47] I.e., a sequential.

> **Definition 4.2.** *Let K be a simplicial complex. A **2-parameter filtration** of K is a collection of subcomplexes $K_{j,k} \leq K$ parameterized with $j, k \in \{1, 2, \ldots, m\}$, such that for each $j \in \{1, 2, \ldots, m-1\}$ and for each k the following containments hold (see Figure 8):*
>
> - $K_{j,k} \leq K_{j+1,k}$, and
>
> - $K_{k,j} \leq K_{k,j+1}$.

There are theoretical and practical settings in which multi-parameter filtrations arise naturally. A **multi-parameter persistent homology** is the object obtained by applying the homology to spaces and maps of such a filtration. Unfortunately, there exists no convenient[48] visualtization[49] in this setting. As a result, theoretical treatments of multi-parameter persistent homology typically deal with a multi-dimensional grid of interconnected homology groups, while practical applications of the same object use incomplete information about it such as multi-parameter tables of Betti numbers, restrictions to a 1-parameter filtrations yielding a standard barcode, etc.
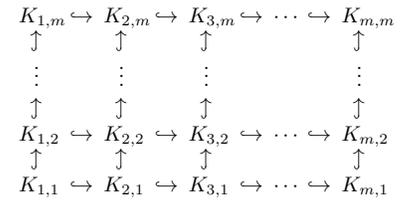
$$\begin{array}{ccccc}
K_{1,m} \hookrightarrow & K_{2,m} \hookrightarrow & K_{3,m} \hookrightarrow & \cdots \hookrightarrow & K_{m,m} \\
\updownarrow & \updownarrow & \updownarrow & & \updownarrow \\
\vdots & \vdots & \vdots & & \vdots \\
\updownarrow & \updownarrow & \updownarrow & & \updownarrow \\
K_{1,2} \hookrightarrow & K_{2,2} \hookrightarrow & K_{3,2} \hookrightarrow & \cdots \hookrightarrow & K_{m,2} \\
\updownarrow & \updownarrow & \updownarrow & & \updownarrow \\
K_{1,1} \hookrightarrow & K_{2,1} \hookrightarrow & K_{3,1} \hookrightarrow & \cdots \hookrightarrow & K_{m,1}
\end{array}$$

Figure 8: A scheme of a 2-parameter filtration.

[48] While a 1-parameter persistent homology "decomposes" into simple pieces called bars (we will explain this statement in detail in the next chapter), the pieces of a multi-parameter persistent homology can be quite complicated and not easily visualized or encoded.

[49] Such as multi-dimensional barcode or persistence diagram.