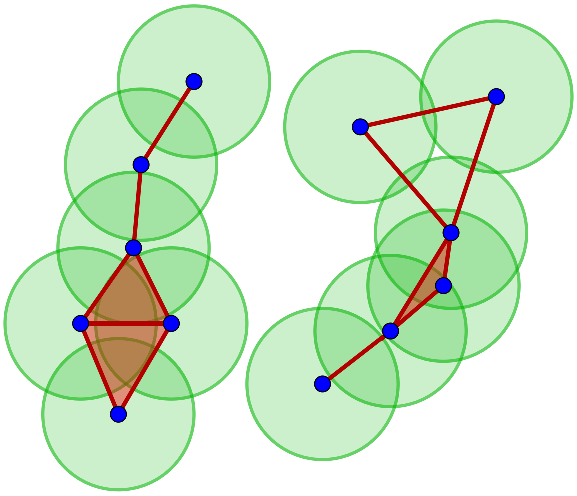
Topological Data Analysis



Motiejus Valiunas

IM UWr, Summer Semester 2022/23

Contents

0

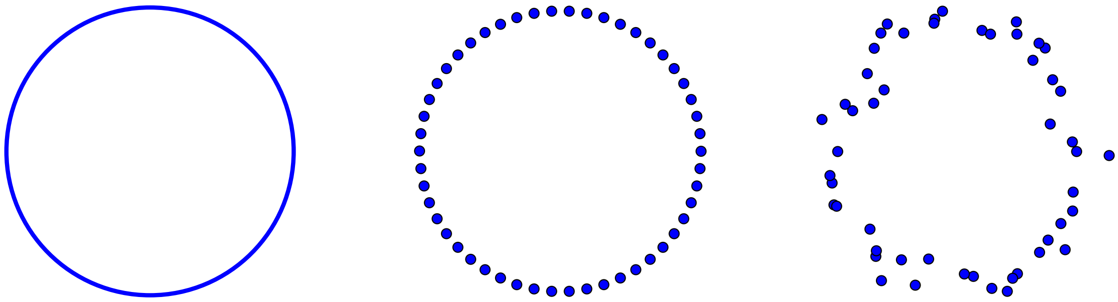
Introduction

Topological Data Analysis (TDA) is a rapidly growing field in mathematics, providing a powerful set of tools for analysing the structure of complex data sets. In the era of big data, it has become increasingly important to be able to extract meaningful information from large and often noisy datasets, and TDA provides a unique approach to this challenge.

TDA is uses mathematical concepts from the fields of topology and geometry to study the shape and structure of data. By focusing on the relationships between data points, rather than just the individual points themselves, TDA is able to capture the underlying patterns in a dataset. This makes it ideal for analysing complex and high-dimensional data sets, where traditional methods may fail.

In this course, our aim is to introduce the main and most well-established method in TDA, namely *persistent homology*. This method works by computing topological features of data at different scales, and then representing the results in a unified way.

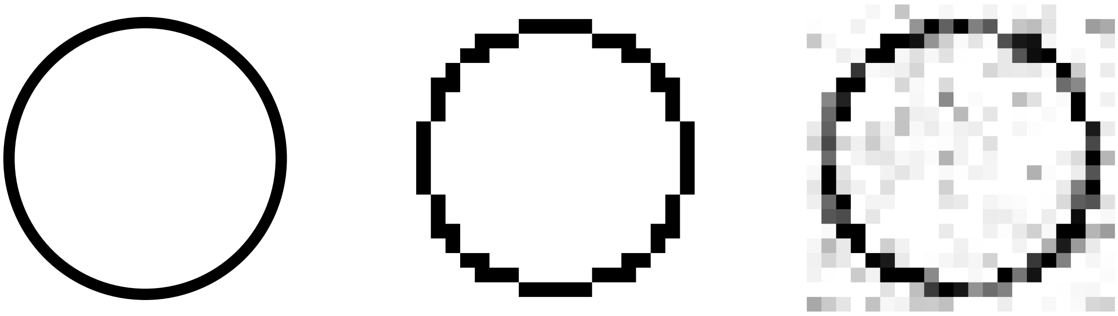
For instance, suppose we have a circle (Figures 1a and 2a). In order to use data analysis, such a circle can be represented in multiple ways. We will concentrate on two such representations: taking a uniform sample of points to form a point cloud (Figure 1b) or representing the data by a picture (Figure 2b). In real life, however, such an "ideal" representation is unlikely and we will usually have to deal with some amount of noise in our data (Figures 1c and 2c).



(a) A circle. (b) Also a circle. (c) Also a circle?

Figure 1: Topological analysis of point clouds.

Our first challenge comes from the fact that we may wish to analyse data coming from different types of sources: point clouds, images, text, etc. We will therefore first need to transform our data into a mathematical object allowing uniform study of its shape. The



(a) A circle. (b) Also a circle. (c) Also a circle?

Figure 2: Topological analysis of pictures.

answer to this is given by the concept of a *simplicial complex*, which can be viewed both geometrically (a shape glued together from Euclidean polyhedra) and combinatorially (which will aid computations). It turns out that a single simplicial complex will not be enough to encode geometric and topological information of the data, and we will need a filtration—an increasing sequence of simplicial complexes.

Question 1. *How do we describe our data in a unified way using a mathematical object?*

Answer. We use simplicial complexes and their filtrations: see Chapter 1.

Once we have this mathematical framework, we will delve into studying the shape of simplicial complexes. This can be done algebraically, using the concept of *homology*—a collection of invariants associated to simplicial complex, one for each integer (*dimension*) $n \geq 0$. Roughly speaking, homology counts connected components (dimension 0), "holes" (dimension 1), "voids" (dimension 2), etc.

Question 2. *How do we count geometric and topological features, such as connected components, "holes" or "voids", in our data?*

Answer. We use homology: see Chapter 2.

Finally, we string all these homology invariants together into a single algebraic object—*persistence module*—associated to a filtration of simplicial complexes. This will allow us to see which structural features we see are "persistent" (and therefore likely represent significant properties of our data), and which are not (hence likely are a result of noise). In order to do that, we will use the fact that persistence modules have been classified, allowing us to represent any such object graphically, using a "persistence diagram" or a "barcode"

Question 3. *How do we decide whether a feature appearing in our data is statistically significant or just a consequence of noise?*

Answer. We use persistence: see Chapter 3.

Filtrations of simplicial complexes

1.1 Simplicial complexes

Definition (simplex)**.** Let $m \geq n \geq 0$, and let $V \subseteq \mathbb{R}^m$ be a finite subset with $|V| = n+1$, so that $V = \{v_0, \ldots, v_n\}$. Suppose that V is *affinely independent*: that is, the n vectors $\{v_i - v_0 \mid 1 \leq i \leq n\}$ are linearly independent.

- The *convex hull* of V is the set $\sigma(V) := \{\sum_{i=0}^n \alpha_i v_i \mid \alpha_i \in [0,1], \sum_{i=0}^n \alpha_i = 1\}$. We call $\sigma(V)$ a *simplex* (or, more precisely, an *n-simplex*).
- Given a non-empty subset $W \subseteq V$, we say that the simplex $\sigma(W)$ is a *face* of $\sigma(V)$ (or, more precisely, a *k-face* of $\sigma(V)$, where $k = |W| - 1$).

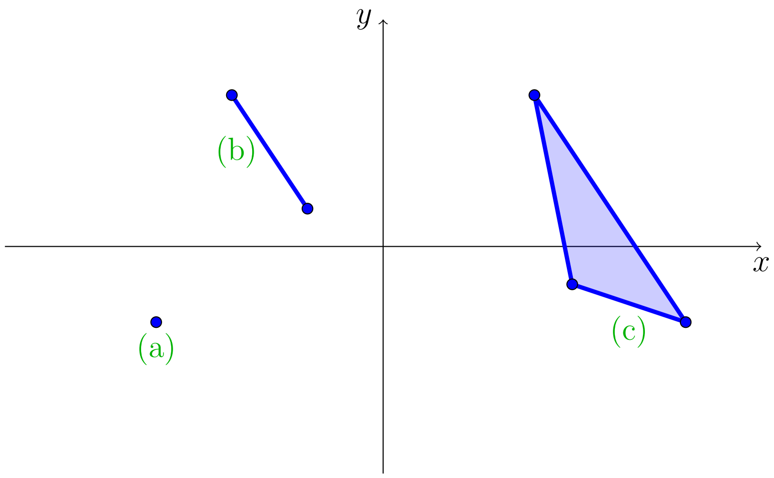


Figure 3: Some simplices in $\mathbb{R}^2$: (a) a 0-simplex, or *vertex*; (b) a 1-simplex, or *edge*, which has a 1-face and two 0-faces; (c) a 2-simplex, or *triangle*, which has a 2-face, three 1-faces and three 0-faces. Note that a triple $\{u, v, w\}$ is affinely independent (and so forms vertices of a triangle) if and only if u, v and w do not lie on the same line.

Definition (geometric simplicial complex)**.** Let $m \geq 0$, and let S be a finite collection of simplices in $\mathbb{R}^m$. We say that S is a *geometric simplicial complex* if

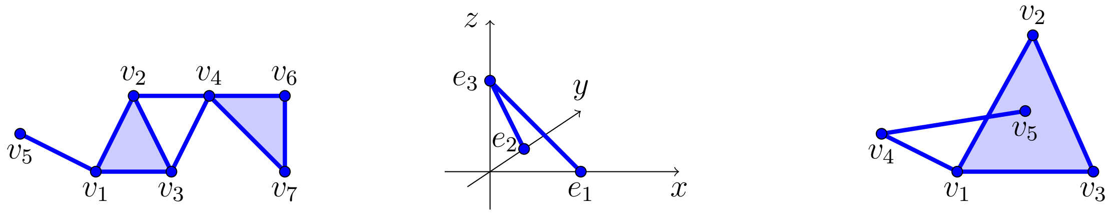
- for any $\sigma \in S$, any face of σ also belongs to S; and
- for any two simplices $\sigma, \tau \in S$, their intersection of $\sigma \cap \tau$ is either empty or a face of both σ and τ.

The *dimension* of S is the largest $n \geq 0$ such that S contains an n-simplex. The *vertex set* $V(S)$ of S is the set of $v \in \mathbb{R}^m$ such that $\{v\}$ is a 0-simplex in S.

Definition (abstract simplicial complex)**.** Let V be a finite set. A collection K of non-empty subsets of V is called an *abstract simplicial complex* if

- $\{v\} \in K$ for all $v \in V$; and
- whenever $W \in K$ and $\varnothing \neq W' \subseteq W$ we also have $W' \in K$.

Elements of K are called *abstract simplices*; more precisely, if $W \in K$ and $|W| = n+1$, then we call W an *abstract n-simplex*. The *dimension* of K is the largest $n \in \mathbb{N}$ such that K contains an abstract n-simplex, and we call $V = V(K)$ the *vertex set* of K. For brevity, we will also write $v_0 v_1 \cdots v_n$ for an abstract simplex $\{v_0, v_1, \ldots, v_n\} \in K$ (in particular, we will not distinguish between a vertex $v \in V(K)$ and an abstract 0-simplex $\{v\} \in K$).



(a) a two-dimensional simplicial complex with 2 triangles, 9 edges and 7 vertices;

(b) the standard geometric realisation of the complex $K = \{\{1\}, \{2\}, \{3\}, \{1,3\}, \{2,3\}\}$;

(c) not a simplicial complex: $\sigma(\{v_5\}) \cap \tau$ is not a face of τ, where $\tau = \sigma(\{v_1, v_2, v_3\})$.

Figure 4: Some geometric simplicial complexes (and something that is not one).

Given a geometric simplicial complex S, we construct an abstract simplicial complex S° as follows: we let $V(S^\circ) = V(S)$, and we set $S^\circ = \{W \subseteq V(S) \mid \sigma(W) \in S\}$. Conversely, given an abstract simplicial complex K, its *geometric realisation* is a geometric simplicial complex $\{\sigma(\{\iota(v_0), \ldots, \iota(v_n)\}) \mid v_0 \cdots v_n \in K\}$, where $\iota \colon V(K) \to \mathbb{R}^m$ is an injective function sending abstract simplices to affinely independent subsets. More explicitly, the *standard geometric realisation* $|K|$ of K can be constructed as follows: given that $V(K) = \{v_1, \ldots, v_m\}$, we set $|K| = \{\sigma(\{e_{i_0}, \ldots, e_{i_n}\}) \mid v_{i_0} \cdots v_{i_n} \in K\}$, where $(e_1, \ldots, e_m)$ is the standard basis of $\mathbb{R}^m$ (so that e_i has i-th coordinate equal to 1 and other coordinates equal to 0).

The operations $S \mapsto S^\circ$ and $K \mapsto |K|$ are in a sense inverse to each other. In particular, if K is an abstract simplicial complex then there is a bijection $V(K) \to V(|K|^\circ)$ sending abstract simplices of K to abstract simplices of $|K|^\circ$; and conversely, given a geometric simplicial complex S there exists a bijection $V(S) \to V(|S^\circ|)$ inducing a correspondence between simplices of S and simplices of $|S^\circ|$.

We will therefore identify abstract and geometric simplicial complexes from now on, and will refer to them as simply *simpicial complexes*. As a general rule, we will use geometric simplicial complexes when drawing pictures and abstract simplicial complexes in definitions and computations.

Example. If X is the geometric simplicial complex displayed in Figure 4a, then

$$X^\circ = \{v_1, v_2, \ldots, v_7, v_1v_2, v_1v_3, v_1v_5, v_2v_3, v_2v_4, v_3v_4, v_4v_6, v_4v_7, v_6v_7, v_1v_2v_3, v_4v_6v_7\}.$$

Finally, we will sometimes need to consider simplicial complexes as metric spaces, motivating the following definition.

Definition (underlying space)**.** Let S be a geometric simplicial complex. The *underlying space* of S is a metric space (X, d), where $X = \bigcup_{\sigma \in S} \sigma \subseteq \mathbb{R}^m$, and given $x, y \in X$ we set $d(x, y)$ to be the length of the shortest path $P \subseteq X$ from x to y. Given an abstract simplicial complex K, the *underlying space* of K is the underlying space of $|K|$.

1.2 Filtrations

Definition (simplicial embedding)**.** Let K and K' be (abstract) simplicial complexes. We say an injective function $f \colon V(K) \to V(K')$ is a *simplicial embedding* if we have $\{f(v_0), \ldots, f(v_n)\} \in K'$ for all $v_0 \cdots v_n \in K$. If this is the case, then f induces an injective function $\overline{f} \colon K \to K'$ given by $\overline{f}(v_0 \cdots v_n) = f(v_0) \cdots f(v_n)$; with a slight abuse of terminology, we will call $\overline{f}$ a simplicial embedding as well, and will not distinguish between f and $\overline{f}$.

Definition (filtration)**.** A *filtration of simplicial complexes* is a collection $\{K_i \mid i \in \mathbb{N}\}$ of simplicial complexes together with a collection $\{f_i \colon K_i \to K_{i+1} \mid i \in \mathbb{N}\}$ of simplicial embeddings. We write $(\{K_i\}, \{f_i\})$ for such a filtration.

In many cases, the simplicial complexes K_i appearing in a filtration $(\{K_i\}, \{f_i\})$ will be *subcomplexes* of another simplicial complex K: that is, we have $K_i \subseteq K$ for all $i \in \mathbb{N}$, with each map f_i being simply the inclusion of K_i into K_{i+1}. If this is the case, we simply write $(\{K_i\}, K)$ for such a filtration.

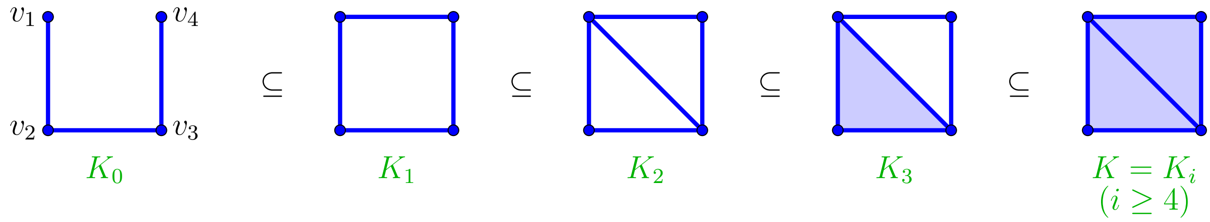


Figure 5: A filtration $(\{K_i\}, K)$ of simplicial complexes.

1.3 Čech and Rips filtrations

Definition (nerve)**.** Let X be a set and $\mathcal{U} = \{U_i, \ldots, U_m\}$ a finite collection of non-empty subsets of X. The *nerve* of $\mathcal{U}$ is the simplicial complex K defined as follows: we let $V(K) = \{v_1, \ldots, v_m\}$, and given any non-empty subset $\sigma = \{v_{i_0}, \ldots, v_{i_n}\} \subseteq V(K)$, we have $\sigma \in K$ if and only if $U_{i_0} \cap \cdots \cap U_{i_n} \neq \varnothing$.

A common construction of nerves involves balls in a metric space. Recall that given a metric space (X, d), a point $x \in X$ and a constant $r \geq 0$, we define the (*closed*) *ball* of radius r around x as the subset $B_r(x) := \{y \in X \mid d(x, y) \leq r\}$ of X.

Definition (Čech filtration)**.** Let (X, d) be a metric space, let $Y \subseteq X$ be a subset with $0 < |Y| < \infty$, and let $\varepsilon > 0$. The *Čech filtration of Y with step ε* is the filtration of simplicial complexes $(\{K_i\}, K)$ defined as follows: we set K to consist of all non-empty subsets of Y (so that K is a $(|Y| - 1)$-simplex together with all of its faces), and for all $i \in \mathbb{N}$ we set K_i to be the nerve of $\{B_{i\varepsilon/2}(y) \mid y \in Y\}$ (in particular, $V(K_i) = V(K) = Y$).

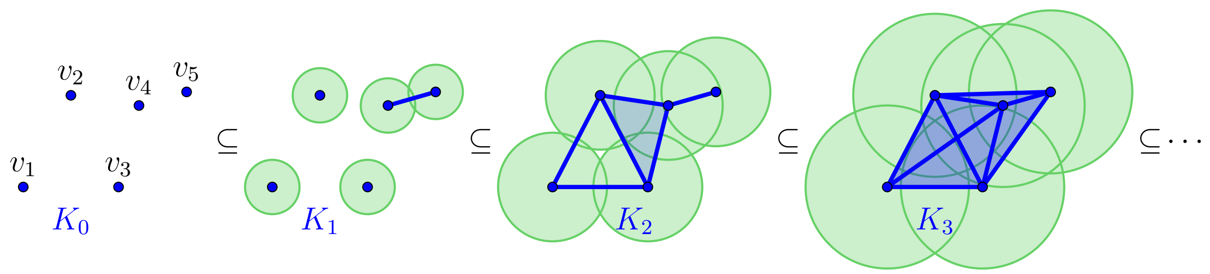
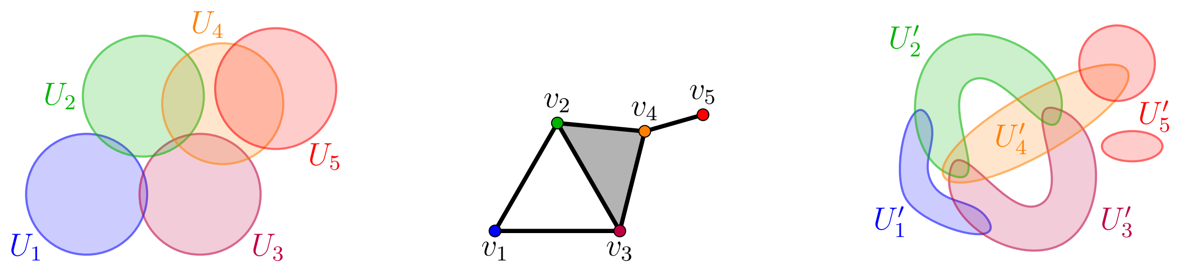


Figure 6: Čech filtration of a subset $\{v_1, v_2, v_3, v_4, v_5\} \subset \mathbb{R}^2$. Here the (abstract) simplicial complexes K_i are visualised in blue, and the balls $B_{i\varepsilon/2}(v_j)$ are drawn in green. In this example, K_3 consists of the 3-simplices $v_1v_2v_3v_4$ and $v_2v_3v_4v_5$ together with their faces, and K_i for $i \geq 4$ consists of the 4-simplex $v_1v_2v_3v_4v_5$ together with its faces. [*Warning:* an abstract simplicial complex of dimension ≥ 3 does not have a geometric realisation in $\mathbb{R}^2$, so e.g. K_3 drawn above is *not* a geometric simplicial complex, only a "visualisation" of an abstract simplicial complex. In general, Čech filtrations usually involve high-dimensional simplicial complexes that cannot be visualised "correctly".]

The theoretical basis for introducing the Čech filtration comes from the following result, using the concept of homotopy equivalence: roughly speaking, two metric spaces are said to be *homotopy equivalent* if one can be continuously transformed into another (without punching holes, making cuts, etc). As a particular interest for us, if two metric spaces are homotopy equivalent then they have the same homology groups (to be defined later in the case of simplicial complexes).

Theorem (Nerve Theorem)**.** *Let $\mathcal{U} = \{U_1, \ldots, U_m\}$ be a collection of non-empty subsets in a metric space (X, d). Suppose that each intersection $U_{i_0} \cap \cdots \cap U_{i_n}$ is either empty or homotopy equivalent to a point. Then the underlying space of the nerve of $\mathcal{U}$ is homotopy equivalent to the union $\bigcup_{i=1}^m U_i$.*

In order to apply this result to Čech filtrations, consider Čech filtrations in the case when (X, d) is the m-dimensional Euclidean space, i.e. $\mathbb{R}^m$ with the "usual" metric. In that case any closed ball B in X is *convex*, i.e. if $x, y \in B$ then the line segment connecting x and y also lies in B. As a consequence, any intersection of balls in $\mathbb{R}^m$ is convex, and any non-empty convex subset of $\mathbb{R}^m$ can be shown to be homotopy equivalent to a point. Therefore, by the Nerve Theorem, each complex K_i appearing in the Čech filtration has its underlying space homotopy equivalent to the union $\bigcup_{y \in Y} B_{i\varepsilon/2}(y)$, and so the Čech filtration recovers the "topology" of these unions.



(a) A collection of convex subsets $U_i \subseteq \mathbb{R}^2$.

(b) The nerve of the collection $\{U_1, U_2, U_3, U_4, U_5\}$.

(c) Another collection of subsets with the same nerve.

Figure 7: An illustration of the Nerve Theorem. The subsets U_i are convex, implying that $\bigcup_i U_i$ is homotopy equivalent to the geometric realisation of the nerve of $\{U_i\}$; this is witnessed by the fact that each of those two spaces has a single connected component with a single "hole" in the middle. On the other hand, $\bigcup_i U'_i$ has two connected components, and one of them has three "holes".

Unfortunately though, computing Čech filtrations in practice is difficult when (X, d) is a Euclidean space, since given a collection of balls $\{B_0, \ldots, B_n\}$ in X, deciding whether $\bigcap_{i=0}^n B_i \neq \varnothing$ is very computationally expensive. Replacing this criterion with (slightly weaker) "$B_i \cap B_j \neq \varnothing$ for all i, j" makes the computations much easier: indeed, if $y, y' \in X$ then we have $B_{i\varepsilon/2}(y) \cap B_{i\varepsilon/2}(y') \neq \varnothing$ if and only if $d(y, y') \leq i\varepsilon$. This motivates the following definition.

Definition (Rips filtration)**.** Let (Y, d) be a metric space with $0 < |Y| < \infty$, and let $\varepsilon > 0$. The *Rips filtration of* (Y, d) *with step* ε is the filtration of simplicial complexes $(\{K_i\}, K)$ defined as follows: we set K to consist of all non-empty subsets of Y (so that K is a $(|Y|-1)$-simplex together with all of its faces), and given any $\sigma \subseteq Y$ and $i \in \mathbb{N}$, we have $\sigma \in K_i$ if and only if $d(v, w) \leq i\varepsilon$ for all $v, w \in \sigma$ (in particular, $V(K_i) = V(K) = Y$).

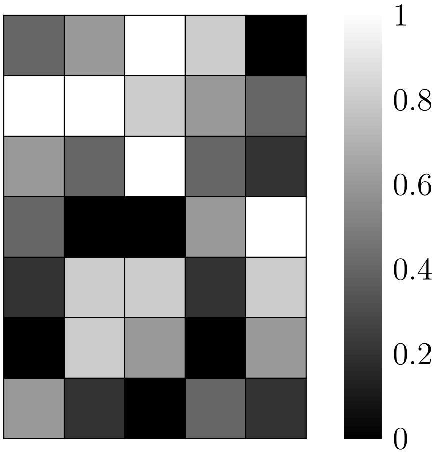
1.4 Morse filtration

Definition (combinatorial Morse function)**.** Let K be a simplicial complex. We say a function $F\colon K \to [0, \infty)$ is a *combinatorial Morse function* if whenever $\sigma \in K$ and $\varnothing \neq \tau \subseteq \sigma$ (that is, τ is a face of σ) we have $F(\tau) \leq F(\sigma)$.

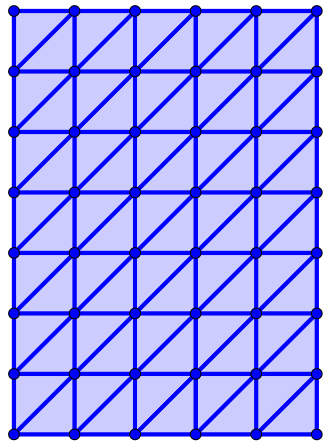
Given a simplicial complex K, any function $\widetilde{F}\colon V(K) \to [0, \infty)$ can be used to construct a combinatorial Morse function $F\colon K \to [0, \infty)$. Indeed, we may set $F(\sigma) := \max\{\widetilde{F}(v_i) \mid 0 \leq i \leq n\}$ for any n-simplex $\sigma = v_0 \cdots v_n \in K$. In this case, we say that F is the combinatorial Morse function *induced* by $\widetilde{F}$. Note, however, that not all combinatorial Morse functions arise this way (as it is possible, for a simplex $\sigma \in K$, that $F(\tau) < F(\sigma)$ for all proper faces $\tau \subsetneq \sigma$).

Definition (Morse filtration)**.** Let $F\colon K \to [0,\infty)$ be a combinatorial Morse function, and let $\varepsilon > 0$. The *Morse filtration of* F *with step* ε is the filtration of simplicial complexes $(\{K_i\}, K)$, where for all $i \in \mathbb{N}$ we set $K_i = \{\sigma \in K \mid F(\sigma) \leq i\varepsilon\}$ (so that $V(K_i)$ consists of all $v \in V(K)$ satisfying $F(v) \leq i\varepsilon$).

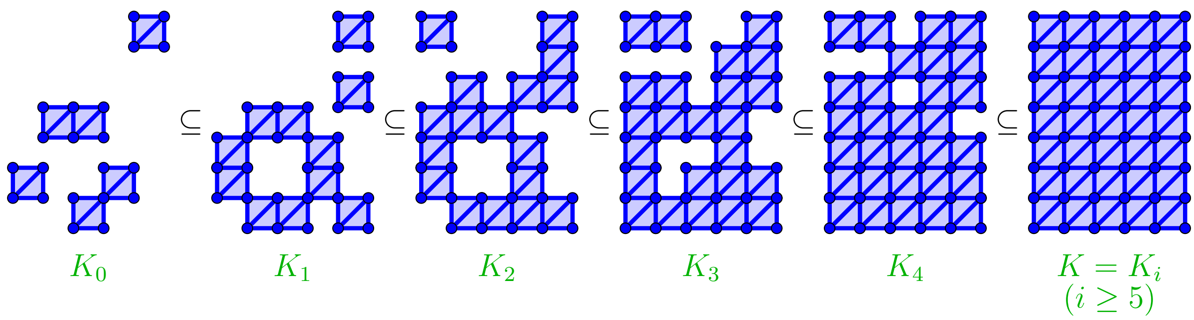
A common application for Morse filtrations is the topological analysis of pictures. In particular, a greyscale picture of width w and height h can be represented by a function $\widehat{F}\colon \{1,\dots,w\}\times\{1,\dots,h\} \to [0,1]$, where $\widehat{F}(i,j)$ represents the colour of the (i,j)-th pixel, with 0 being black and 1 being white. By replacing any pair $(i,j) \in \{1,\dots,w\}\times\{1,\dots,h\}$ with a square $[i-1,i]\times[j-1,j] \subset \mathbb{R}^2$ and subdividing each such square into two triangles, one gets a 2-dimensional simplicial complex K with $2wh$ triangles, and the function $\widehat{F}$ induces a function $F^{(2)}\colon V^{(2)}(K) \to [0,1]$, where $V^{(2)}(K)$ is the set of triangles in K. We can then extend $F^{(2)}$ to a combinatorial Morse function $F\colon K \to [0,\infty)$ defined by setting $F(\tau)$ to be the minimal value of $F^{(2)}(\sigma)$, where σ ranges over all triangles in K such that $\tau \subseteq \sigma$.



(a) A greyscale picture.



(b) The corresponding simplicial complex K.



(c) The Morse filtration of F with step $\varepsilon = 0.2$, where $F\colon K \to [0,\infty)$ is the combinatorial Morse function corresponding to the displayed greyscale picture.

Figure 8: A Morse filtration of the combinatorial Morse function associated to a greyscale picture.

Homology

2.1 Chain complexes

Definition (chain complex)**.** Let $\mathbb{F}$ be a field. A *chain complex* over $\mathbb{F}$ is a collection $\{C_n \mid n \in \mathbb{N}\} \cup \{C_{-1} = 0\}$ of $\mathbb{F}$-vector spaces together with a collection $\{\partial_n \colon C_n \to C_{n-1} \mid n \in \mathbb{N}\}$ of $\mathbb{F}$-linear maps such that $\partial_n \circ \partial_{n+1} \colon C_{n+1} \to C_{n-1}$ is the zero map for all $n \in \mathbb{N}$. We denote such a chain complex by $(C_\bullet, \partial_\bullet)$, and call elements of the C_n *chains*.

Given a simplicial complex K with $V(K) = \{v_1, \ldots, v_m\}$ (so that K has m vertices) and a field $\mathbb{F}$, we construct a chain complex $(C_\bullet(K;\mathbb{F}), \partial_\bullet)$ over $\mathbb{F}$ as follows. For any $n \in \mathbb{N}$, let $C_n(K;\mathbb{F})$ be the $\mathbb{F}$-vector space with basis $\{e_\sigma \mid \sigma \in K \text{ is an } n\text{-simplex}\}$. In order to define the maps $\partial_n \colon C_n(K;\mathbb{F}) \to C_{n-1}(K;\mathbb{F})$, it is then enough to choose $\partial_n(e_\sigma)$ for any $n \in \mathbb{N}$ and any n-simplex $\sigma \in K$. First, we set $\partial_0(e_{v_i}) := 0$ for $1 \leq i \leq m$. Given any $n \geq 1$ and any n-simplex $\sigma = v_{i_0} \cdots v_{i_n} \in K$ with $i_0 < i_1 < \cdots < i_n$, we set

$$\partial_n(e_\sigma) := \sum_{j=0}^{n} (-1)^j e_{\sigma\langle j\rangle} = e_{\sigma\langle 0\rangle} - e_{\sigma\langle 1\rangle} + e_{\sigma\langle 2\rangle} - \cdots + (-1)^n e_{\sigma\langle n\rangle},$$

where $\sigma\langle j\rangle := \sigma \setminus \{v_{i_j}\} = v_{i_0} \cdots v_{i_{j-1}} v_{i_{j+1}} \cdots v_{i_n}$ is an $(n-1)$-face of σ.

Example. Let K be the simplicial complex displayed in Figure 4a, and let $\mathbb{F}$ be any field. The resulting chain complex $(C_\bullet(K;\mathbb{F}), \partial_\bullet)$ then satisfies $C_0(K;\mathbb{F}) \cong \mathbb{F}^7$, $C_1(K;\mathbb{F}) \cong \mathbb{F}^9$, $C_2(K;\mathbb{F}) \cong \mathbb{F}^2$ and $C_n(K;\mathbb{F}) = 0$ for $n \geq 3$. Furthermore, the maps ∂_n can be represented by the following matrices (with respect to the "obvious" bases):

$$\partial_n = 0 \colon 0 \to 0 \text{ for } n \geq 4, \qquad \partial_3 = 0 \colon 0 \to \mathbb{F}^2,$$

$$[\partial_2] = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 1 \end{pmatrix}, \qquad [\partial_1] = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

$$\text{and} \qquad \partial_0 = 0 \colon \mathbb{F}^7 \to 0.$$

We now show that, in general, this construction actually defines a chain complex.

Proposition. *$(C_\bullet(K;\mathbb{F}),\partial_\bullet)$ is a chain complex.*

Proof. We need to show that $\partial_n \circ \partial_{n+1}$ is the zero map for all $n \in \mathbb{N}$. In order to do that, it is enough to show that $(\partial_n \circ \partial_{n+1})(e_\sigma) = 0$ for any $(n+1)$-simplex $\sigma \in K$ and any $n \in \mathbb{N}$. Since ∂_0 is the zero map, we may assume that $n \geq 1$; given an $(n+1)$-simplex $\sigma = v_{i_0} \cdots v_{i_{n+1}} \in K$, we then have

$$\begin{aligned}(\partial_n \circ \partial_{n+1})(e_\sigma) &= \partial_n\left(\sum_{j=0}^{n+1}(-1)^j e_{\sigma\langle j\rangle}\right) = \sum_{j=0}^{n+1}(-1)^j \partial_n(e_{\sigma\langle j\rangle}) = \sum_{j=0}^{n+1}(-1)^j \sum_{k=0}^{n}(-1)^k e_{\sigma\langle j\rangle\langle k\rangle} \\ &= \sum_{j=0}^{n+1}\sum_{k=0}^{n}(-1)^{j+k} e_{\sigma\langle j\rangle\langle k\rangle}.\end{aligned}$$

Now the terms on the right hand side are all of the form e_τ, where τ is an $(n-1)$-face of σ. Moreover, given such a face $\tau = \sigma \setminus \{v_{i_{j'}}, v_{i_{k'}}\}$, where $j' < k'$, the simplex e_τ appears in the sum in two places: once as a term in $\partial_n(e_{\sigma\langle j'\rangle})$, and once as a term in $\partial_n(e_{\sigma\langle k'\rangle})$. Since we have $j' < k'$, the vertex $v_{k'}$ appears as the $(k'-1)$-st vertex of $\sigma\langle j'\rangle$ in the chosen ordering, whereas $v_{j'}$ is still the j'-th vertex of $\sigma\langle k'\rangle$. Therefore, we have $\tau = \sigma\langle j'\rangle\langle k'-1\rangle = \sigma\langle k'\rangle\langle j'\rangle$, implying that the term $(-1)^{j'+(k'-1)} e_{\sigma\langle j'\rangle\langle k'-1\rangle} = -(-1)^{j'+k'} e_\tau$ cancels out with the term $(-1)^{k'+j'} e_{\sigma\langle k'\rangle\langle j'\rangle} = (-1)^{j'+k'} e_\tau$, and so the total contribution of e_τ to the sum is zero, as required. □

Example. Given a 2-simplex $v_i v_j v_k \in K$ with $i < j < k$, we have

$$\partial_2(e_{v_i v_j v_k}) = e_{v_j v_k} - e_{v_i v_k} + e_{v_i v_j} \qquad \text{and} \qquad \partial_1(e_{v_i v_j}) = e_{v_j} - e_{v_i}.$$

We may compute that

$$\begin{aligned}(\partial_1 \circ \partial_2)(e_{v_i v_j v_k}) &= \partial_1(e_{v_j v_k}) - \partial_1(e_{v_i v_k}) + \partial_1(e_{v_i v_j}) \\ &= (e_{v_k} - e_{v_j}) - (e_{v_k} - e_{v_i}) + (e_{v_j} - e_{v_i}) = 0.\end{aligned}$$

Finally, note that the definition of $(C_\bullet(K;\mathbb{F}),\partial_\bullet)$ depends on the order we chose for listing the vertices in $V(K)$. However, this choice is inessential in the end: as the next result shows, a different ordering will give the same chain complex up to isomorphism.

Lemma. *Let $V(K) = \{v_1,\dots,v_m\} = \{v_1',\dots,v_m'\}$, and let $(C_\bullet(K;\mathbb{F}),\partial_\bullet)$, $(C_\bullet(K;\mathbb{F}),\partial_\bullet')$ be the corresponding chain complexes. Then for each $n \in \mathbb{N}$ there exists an isomorphism $\iota_n\colon C_n(K;\mathbb{F}) \to C_n(K;\mathbb{F})$ such that $\partial_n' = \iota_{n-1} \circ \partial_n \circ \iota_n^{-1}$.*

Proof. We first claim that one can obtain the sequence $(v_1,\dots,v_m)$ from the sequence $(v_1',\dots,v_m')$ by succesively swapping pairs of consecutive vertices. Indeed, we first "put the vertex v_m in place": if $j \in \{1,\dots,m\}$ is such that $v_j' = v_m$, then we can swap the j-th vertex with the $(j+1)$-st ($v_j' \leftrightarrow v_{j+1}'$), then the $(j+1)$-st with the $(j+2)$-nd ($v_j' \leftrightarrow v_{j+2}'$), and continue this way until we swap the $(m-1)$-st vertex with the m-th ($v_j' \leftrightarrow v_m'$). Repeating this procedure for v_{m-1}, then v_{m-2}, and so on down to v_2, we obtain the sequence $(v_1,\dots,v_m)$, as required.

Therefore, it is enough to show that the isomorphisms ι_n exist when $(v_1, \ldots, v_m)$ can be obtained from $(v'_1, \ldots, v'_m)$ in a single step of this swapping process. That is, we may assume without loss of generality that $v'_j = v_{j+1}$, $v'_{j+1} = v_j$ (for some $j \in \{1, \ldots, m-1\}$) and $v'_i = v_i$ for all $i \notin \{j, j+1\}$. In this case, define ι_n by setting $\iota_n(e_\sigma) = -e_\sigma$ if $\{j, j+1\} \subseteq \sigma$ and $\iota_n(e_\sigma) = e_\sigma$ otherwise (where $\sigma \in K$ is an n-simplex).

It is now enough to show that $(\partial'_n \circ \iota_n)(e_\sigma) = (\iota_{n-1} \circ \partial_n)(e_\sigma)$ for any n-simplex $\sigma \in K$. If $\{j, j+1\} \nsubseteq \sigma$ then we have

$$(\iota_{n-1} \circ \partial_n)(e_\sigma) = \sum_{k=0}^{n} (-1)^k \iota_{n-1}(e_{\sigma\langle k\rangle}) = \sum_{k=0}^{n} (-1)^k e_{\sigma\langle k\rangle} = \partial'_n(e_\sigma) = (\partial'_n \circ \iota_n)(e_\sigma),$$

as required. On the other hand, if $\sigma = v_{i_0} \cdots v_{i_n}$ where $i_0 < \ldots < i_n$ and we have $i_k = j$ and $i_{k+1} = j+1$ for some $k \in \{0, \ldots, n-1\}$, it then follows that $\iota_{n-1}(e_{\sigma\langle k'\rangle}) = e_{\sigma\langle k'\rangle}$ if $k' \in \{k, k+1\}$ and $\iota_{n-1}(e_{\sigma\langle k'\rangle}) = -e_{\sigma\langle k'\rangle}$ otherwise, and therefore

$$\begin{aligned}
(\partial'_n \circ \iota_n)(e_\sigma) &= \partial'_n(-e_\sigma) = -\partial'_n(e_\sigma) \\
&= -\left(\sum_{k'=0}^{k-1} (-1)^{k'} e_{\sigma\langle k'\rangle} + (-1)^k e_{\sigma\langle k+1\rangle} + (-1)^{k+1} e_{\sigma\langle k\rangle} + \sum_{k'=k+2}^{n} (-1)^{k'} e_{\sigma\langle k'\rangle} \right) \\
&= \sum_{k'=0}^{k-1} (-1)^{k'} (-e_{\sigma\langle k'\rangle}) + (-1)^{k+1} e_{\sigma\langle k+1\rangle} + (-1)^k e_{\sigma\langle k\rangle} + \sum_{k'=k+2}^{n} (-1)^{k'} (-e_{\sigma\langle k'\rangle}) \\
&= \sum_{k'=0}^{n} (-1)^{k'} \iota_{n-1}(e_{\sigma\langle k'\rangle}) = (\iota_{n-1} \circ \partial_n)(e_\sigma),
\end{aligned}$$

as required. □

2.2 Homology groups

Now suppose we are given a chain complex $(C_\bullet, \partial_\bullet)$. The condition $\partial_n \circ \partial_{n+1} = 0$ means precisely that the image of ∂_{n+1}, i.e. the subspace $\operatorname{im}(\partial_{n+1}) = \{\partial_{n+1}(x) \mid x \in C_{n+1}\} \leq C_n$, is a subspace of the kernel of ∂_n, i.e. the subspace $\ker(\partial_n) = \{y \in C_n \mid \partial_n(y) = 0\} \leq C_n$. This motivates the following definition.

Definition (homology groups)**.** Let $\mathbb{F}$ be a field.

- For a chain complex $(C_\bullet, \partial_\bullet)$ over $\mathbb{F}$ and $n \in \mathbb{N}$, the n-th *homology group* of $(C_\bullet, \partial_\bullet)$ is the quotient $\mathbb{F}$-vector space $H_n(C_\bullet, \partial_\bullet) := \ker(\partial_n) / \operatorname{im}(\partial_{n+1})$.

- Given a simplicial complex K, the n-th *homology group* of K over $\mathbb{F}$ is the $\mathbb{F}$-vector space $H_n(K; \mathbb{F}) := H_n(C_\bullet(K; \mathbb{F}), \partial_\bullet)$, where $(C_\bullet(K; \mathbb{F}), \partial_\bullet)$ is the chain complex constructed above.

Informally, we can think of a homology group $H_n(K; \mathbb{F})$ as counting the n-dimensional "holes" in (the underlying space of) K. Specifically, $\dim_{\mathbb{F}} H_0(K; \mathbb{F})$ is the number of connected components in K, $\dim_{\mathbb{F}} H_1(K;\mathbb{F})$ is the number of (1-dimensional) "holes" (such as the empty area inside a circle), and $\dim_{\mathbb{F}} H_2(K;\mathbb{F})$ is the number of 2-dimensional "voids" (such as the empty space inside a sphere).

Example. Let K be the simplicial complex displayed in Figure 4a, and let $\mathbb{F}$ be any field. From the computation of the complex $(C_\bullet(K;\mathbb{F}),\partial_\bullet)$ above, it follows that we have

$$\begin{aligned}
\ker(\partial_n) &= 0 && \text{and } \operatorname{im}(\partial_n) = 0 \qquad \text{for } n \geq 3;\\
\ker(\partial_2) &= 0 && \text{and } \operatorname{im}(\partial_2) = \operatorname{span}\{\partial_2(e_{v_1v_2v_3}), \partial_2(e_{v_4v_6v_7})\} \cong \mathbb{F}^2;\\
\ker(\partial_1) &= \operatorname{span}\{\partial_2(e_{v_1v_2v_3}), \partial_2(e_{v_4v_6v_7}), \\
&\qquad e_{v_2v_3} - e_{v_2v_4} + e_{v_3v_4}\} \cong \mathbb{F}^3 && \text{and } \operatorname{im}(\partial_1) = \{\textstyle\sum_{i=1}^7 \alpha_i e_{v_i} \mid \sum_{i=1}^7 \alpha_i = 0\} \cong \mathbb{F}^6;\\
\ker(\partial_0) &= C_0(K;\mathbb{F}) \cong \mathbb{F}^7 && \text{and } \operatorname{im}(\partial_0) = 0.
\end{aligned}$$

This implies that $H_0(K;\mathbb{F}) \cong H_1(K;\mathbb{F}) \cong \mathbb{F}$ and $H_n(K;\mathbb{F}) = 0$ for $n \geq 2$.

In the example above, the dimension $\dim_{\mathbb{F}} H_n(K;\mathbb{F})$ is independent of the choice of the field $\mathbb{F}$. However, this is not true in general. Examples demonstrating this involve high-dimensional vector spaces and are therefore hard to compute by hand; nevertheless, a similar behaviour is seen in some lower-dimensional examples of homology of chain complexes, as the following example shows.

Example. Let $\mathbb{F}$ be a field, and define a chain complex $(C_\bullet,\partial_\bullet)$ as follows. Let $C_0 = C_1 = \mathbb{F}^2$ and $C_n = 0$ for $n \geq 2$, let $\partial_n \colon C_n \to C_{n-1}$ be the zero map for $n \neq 1$, and let $\partial_1 \colon \mathbb{F}^2 \to \mathbb{F}^2$ be represented by the matrix

$$[\partial_1] = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

We can then compute that $H_n(C_\bullet,\partial_\bullet) = 0$ for $n \geq 2$, regardless of the choice of the field $\mathbb{F}$. Now if $\mathbb{F}$ is a field of characteristic 2, such as the field $\mathbb{F}_2$ of two elements, then the map ∂_1 has both its kernel and image equal to $\{(x,x) \mid x \in \mathbb{F}\}$, implying that $H_0(C_\bullet,\partial_\bullet) \cong H_1(C_\bullet,\partial_\bullet) \cong \mathbb{F}$. However, if $\mathbb{F}$ is any other field, e.g. $\mathbb{F} = \mathbb{Q},\mathbb{R}$ or $\mathbb{C}$, then the map ∂_1 is an isomorphism and we have $H_0(C_\bullet,\partial_\bullet) = H_1(C_\bullet,\partial_\bullet) = 0$.

An important feature of homology is that it is "homotopy-invariant": that is, if two (underlying spaces of) simplicial complexes K and K' are homotopy equivalent, then the homology groups $H_n(K;\mathbb{F})$ and $H_n(K';\mathbb{F})$ are isomorphic, that is, have the same dimension (for fixed $\mathbb{F}$ and n).

2.3 Euler characteristic

Definition. Let K be a simplicial complex, and for each $n \in \mathbb{N}$ denote by $c_n(K)$ the number of n-simplices of K. Then the *Euler characteristic* of K, denoted $\chi(K)$, is the integer $\chi(K) := \sum_{n=0}^{\infty} (-1)^n c_n(K)$.

The following result shows that the Euler characteristic can be computed only from the homology groups of a simplicial complex K. In particular, if the simplicial complexes K and K' are homotopy equivalent, then $\chi(K) = \chi(K')$.

Lemma. *Let K be a simplicial complex, let $\mathbb{F}$ be a field, and, for each $n \in \mathbb{N}$, define $\beta_n(K;\mathbb{F}) := \dim_{\mathbb{F}} H_n(K;\mathbb{F})$ (called the n-th* Betti number *of K over $\mathbb{F}$). Then we have $\chi(K) = \sum_{n=0}^{\infty}(-1)^n \beta_n(K;\mathbb{F})$.*

Proof. Let $(C_\bullet(K;\mathbb{F}), \partial_\bullet)$ be the corresponding chain complex, and write k_n and d_n for $\dim_{\mathbb{F}} \ker(\partial_n)$ and $\dim_{\mathbb{F}} \operatorname{im}(\partial_n)$, respectively. For each $n \in \mathbb{N}$, we have $\beta_n(K;\mathbb{F}) = k_n - d_{n+1}$ by the definition of homology groups, whereas $k_n + d_n = \dim_{\mathbb{F}} C_n(K;\mathbb{F}) = c_n(K)$ by the Rank-Nullity Theorem. Let $N \in \mathbb{N}$ be the dimension of K, and note that $\partial_0 = 0$, $\partial_{N+1} = 0$, and $c_n(K) = \beta_n(K;\mathbb{F}) = 0$ for $n > N$. We then have

$$\begin{aligned}
\sum_{n=0}^{\infty}(-1)^n \beta_n(K;\mathbb{F}) &= \sum_{n=0}^{N}(-1)^n \left[k_n - d_{n+1}\right] = \sum_{n=0}^{N}(-1)^n k_n + \sum_{n=1}^{N+1}(-1)^n d_n \\
&= (-1)^{N+1} d_{N+1} - d_0 + \sum_{n=0}^{N}(-1)^n \left[k_n + d_n\right] \\
&= (-1)^{N+1} 0 - 0 + \sum_{n=0}^{N}(-1)^n c_n(K) = \chi(K),
\end{aligned}$$

as required. □

Example. Let K be the simplicial complex displayed in Figure 4a, and let $\mathbb{F}$ be any field. We have $c_0(K) = 7$, $c_1(K) = 9$, $c_2(K) = 2$ and $c_n(K) = 0$ for $n \geq 3$, implying that $\chi(K) = 7 - 9 + 2 = 0$. Alternatively, it follows from the computation of homology above that $\beta_0(K;\mathbb{F}) = \beta_1(K;\mathbb{F}) = 1$ and $\beta_n(K;\mathbb{F}) = 0$ for $n \geq 2$, implying again that $\chi(K) = 1 - 1 = 0$.

2.4 Functoriality

The idea of persistent homology (to be discussed later) is to start with a filtration of simplicial complexes $(\{K_i\}, \{f_i\})$, and combine the homology groups $\{H_n(K_i) \mid i \in \mathbb{N}\}$ into a single object. In order to do that, we need to know that simplicial embeddings $f_i \colon K_i \to K_{i+1}$ induce maps between the corresponding homology groups.

Proposition. *Let $f \colon K \to K'$ be a simplicial embedding between simplicial complexes K and K', and let $\mathbb{F}$ be a field. Then there is a "canonical" induced map $f_* \colon H_n(K;\mathbb{F}) \to H_n(K';\mathbb{F})$ for every $n \in \mathbb{N}$.*

In the notation of the Proposition, we say that f *induces f_* in homology.*

Proof. Let $(C_\bullet(K;\mathbb{F}), \partial_\bullet^K)$ and $(C_\bullet(K';\mathbb{F}), \partial_\bullet^{K'})$ be the chain complexes defined using labellings $V(K) = \{v_1, \ldots, v_m\}$ and $V(K') = \{v'_1, \ldots, v'_{m'}\}$. Since the chain complex

$(C_\bullet(K;\mathbb{F}), \partial_\bullet^K)$ does not depend on the ordering chosen in an essential way, we may assume that the vertices of K are ordered so that if $f(v_j) = v_k$ and $f(v_{j'}) = v_{k'}$, then $j < j'$ if and only if $k < k'$. Then for each $n \in \mathbb{N}$, the map f induces an $\mathbb{F}$-linear map $f_n\colon C_n(K;\mathbb{F}) \to C_n(K';\mathbb{F})$ given by setting $f_n(e_\sigma) := e_{f(\sigma)}$ for any n-simplex $\sigma \in K$. We also set $f_{-1}\colon 0 \to 0$ to be the zero map.

We first claim that $f_{n-1} \circ \partial_n^K = \partial_n^{K'} \circ f_n$ for all $n \in \mathbb{N}$. For $n = 0$ both of these maps are the zero map, so we can assume that $n \geq 1$. It is then enough to show that $(f_{n-1} \circ \partial_n^K)(e_\sigma) = (\partial_n^{K'} \circ f_n)(e_\sigma)$ for any n-simplex $\sigma \in K$. Note that, since the order on $V(K)$ is chosen to be consistent with the order on $V(K')$, we have $f(\sigma\langle j\rangle) = f(\sigma)\langle j\rangle$ for all j. It follows that

$$\begin{aligned}(f_{n-1} \circ \partial_n^K)(e_\sigma) &= f_{n-1}\left(\sum_{j=0}^{n}(-1)^j e_{\sigma\langle j\rangle}\right) = \sum_{j=0}^{n}(-1)^j e_{f(\sigma\langle j\rangle)} \\ &= \sum_{j=0}^{n}(-1)^j e_{f(\sigma)\langle j\rangle} = \partial_n^{K'}(e_{f(\sigma)}) = (\partial_n^{K'} \circ f_n)(e_\sigma),\end{aligned}$$

as claimed.

Now in order to show that f_n induces a map $f_*\colon H_n(K;\mathbb{F}) \to H_n(K';\mathbb{F})$, it is enough to show that $f_n(\operatorname{im}(\partial_{n+1}^K)) \subseteq \operatorname{im}(\partial_{n+1}^{K'})$ (so that f_n induces a well-defined quotient map $f'_n\colon C_n(K;\mathbb{F})/\operatorname{im}(\partial_{n+1}^K) \to C_n(K';\mathbb{F})/\operatorname{im}(\partial_{n+1}^{K'})$), and that $f_n(\ker(\partial_n^K)) \subseteq \ker(\partial_n^{K'})$ (so that the restriction f_* of f'_n to $H_n(K;\mathbb{F})$ has image in $H_n(K',\mathbb{F})$. For the former, note that if $c \in \operatorname{im}(\partial_{n+1}^K)$ then $c = \partial_{n+1}^K(c')$ for some $c' \in C_{n+1}(K;\mathbb{F})$ and therefore

$$f_n(c) = (f_n \circ \partial_{n+1}^K)(c') = \partial_{n+1}^{K'}(f_{n+1}(c')) \in \operatorname{im}(\partial_{n+1}^{K'}).$$

For the latter, note that if $c \in \ker(\partial_n^K)$ then

$$\partial_n^{K'}(f_n(c)) = (f_{n-1} \circ \partial_n^K)(c) = f_{n-1}(0) = 0$$

and therefore $f_n(c) \in \ker(\partial_n^{K'})$. This proves that the map $f_*\colon H_n(K;\mathbb{F}) \to H_n(K';\mathbb{F})$ is well-defined, as required. □

3

Persistent homology

3.1 Persistence modules

We first introduce the notion of modules; roughly speaking, an R-module is a generalisation of an R-vector space for commutative rings R that are not necessarily fields.

Definition (module)**.** Let R be a commutative ring with 1.

- An *R-module* M is an abelian group $(M, +)$ together with an operation $R \times M \to M$ sending $(r, m) \mapsto r \cdot m$ such that $r \cdot (x + y) = r \cdot x + r \cdot y$, $(r + s) \cdot x = r \cdot x + s \cdot x$, $(rs) \cdot x = r \cdot (s \cdot x)$ and $1 \cdot x = x$ for all $r, s \in R$ and $x, y \in M$.
- A subset $\{x_1, \ldots, x_m\} \subseteq M$ is called a (*finite*) *generating set* if every element $y \in M$ can be expressed as $y = \sum_{i=1}^{m} r_i \cdot x_i$ for some $r_1, \ldots, r_m \in R$, and we say M is *finitely generated* if it has a finite generating set.

Example. Let R be a commutative ring with 1.

- The ring R itself is an R-module. More generally, any ideal $I \trianglelefteq R$ and any R^n (for an integer $n \geq 1$) are R-modules (we call R^n a *free* R-module).
- If R is a field, then an R-module is just an R-vector space, and it is finitely generated as a module if and only if it is finite-dimensional as a vector space.
- A $\mathbb{Z}$-module is just an abelian group. It can be finitely generated (e.g. $\mathbb{Z}^n$, finite abelian groups) or not (e.g. $\mathbb{Q}$).
- If $R = \mathbb{Z}/n\mathbb{Z}$ for some $n \geq 2$, then an R-module is an abelian group in which the order of each element is divisible by n.
- If $R = \mathbb{Z}[i]$ where $i = \sqrt{-1}$, then an R-module is an abelian group M together with an isomorphism $f \colon M \to M$ such that $f(f(y)) = -y$ for all $y \in M$ (in particular, f is defined by setting $f(y) = i \cdot y$).

We will be particularly interested in R-modules when $R = \mathbb{F}[X]$, the polynomial ring over a field $\mathbb{F}$ in one variable X. In this case, we can also define graded modules, as follows:

Definition (graded module)**.** Let $\mathbb{F}$ be a field. An $\mathbb{F}[X]$-module M is said to be *graded* if $M = \bigoplus_{i=0}^{\infty} M_i$ as an $\mathbb{F}$-vector space (for some subspaces $M_0, M_1, \ldots$) and $X \cdot y \in M_{i+1}$ whenever $y \in M_i$.

Now for any field $\mathbb{F}$, given a collection $\{M_i \mid i \in \mathbb{N}\}$ of $\mathbb{F}$-vector spaces and a collection $\{f_i \colon M_i \to M_{i+1} \mid i \in \mathbb{N}\}$ of $\mathbb{F}$-linear maps, one can form a graded $\mathbb{F}[X]$-module M by setting $M := \bigoplus_{i=0}^{\infty} M_i$ (as $\mathbb{F}$-vector spaces) and $X \cdot y := f_i(y)$ whenever $y \in M_i$: this uniquely extends to a map $\mathbb{F}[X] \times M \to M, (P, y) \to P \cdot y$ satisfying the axioms for an $\mathbb{F}[X]$-module. Conversely, any graded $\mathbb{F}[X]$-module gives rise to such collections $\{M_i \mid i \in \mathbb{N}\}$ and $\{f_i \colon M_i \to M_{i+1} \mid i \in \mathbb{N}\}$.

Definition. Let $\mathbb{F}$ be a field, let $(\{K_i\}, \{f_i\})$ be a filtration of simplicial complexes, and let $n \in \mathbb{N}$. The n-th *persistence module* of $(\{K_i\}, \{f_i\})$ over $\mathbb{F}$, denoted by $H_n(\{K_i\}, \{f_i\}; \mathbb{F})$, is a graded $\mathbb{F}[X]$-module equal to $\bigoplus_{i=0}^{\infty} H_n(K_i; \mathbb{F})$ as an $\mathbb{F}$-vector space (with the obvious grading), made into a module by setting $X \cdot y = (f_i)_*(y)$ for $y \in H_n(K_i; \mathbb{F})$, where $(f_i)_* \colon H_n(K_i; \mathbb{F}) \to H_n(K_{i+1}, \mathbb{F})$ is the map induced in homology by $f_i \colon K_i \to K_{i+1}$.

Example. Let $(\{K_i\}, K)$ be the filtration of simplicial complexes displayed in Figure 5. We may then compute that

$$H_1(K_1; \mathbb{F}) = \mathbb{F}(e'_1), \quad H_1(K_2; \mathbb{F}) = \mathbb{F}^2(e'_2, e'_3), \quad H_1(K_3; \mathbb{F}) = \mathbb{F}(e'_4)$$

and $H_1(K_i; \mathbb{F}) = 0$ for $i \notin \{1, 2, 3\}$, where the images of the chains shown in brackets form bases of the corresponding $\mathbb{F}$-vector spaces and we write

$$e'_1 := e_{v_1v_2} + e_{v_2v_3} + e_{v_3v_4} - e_{v_1v_4}, \qquad e'_2 := e_{v_1v_2} + e_{v_2v_3} - e_{v_1v_3}$$
$$\text{and} \qquad e'_3 = e'_4 := e_{v_1v_3} + e_{v_3v_4} - e_{v_1v_4}.$$

The inclusion maps $f_i \colon K_i \to K_{i+1}$ then induce the maps $(f_i)_*$ in homology, where (with respect to the obvious bases) we have $[(f_1)_*] = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $[(f_2)_*] = \begin{pmatrix} 0 & 1 \end{pmatrix}$ and $(f_i)_* = 0$ for $i \notin \{1, 2\}$. The corresponding persistence module $H_1(\{K_i\}, K; \mathbb{F})$ is then isomorphic to $\mathbb{F}^4$ as an $\mathbb{F}$-vector space, with the map $y \mapsto X \cdot y$ represented by the matrix

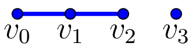
$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

with respect to the basis (e'_1, e'_2, e'_3, e'_4).

Proposition. *Let $(\{K_i\}, K)$ be a filtration of simplicial complexes, let $n \in \mathbb{N}$, and let $\mathbb{F}$ be a field. Then the persistence module $H_n(\{K_i\}, K; \mathbb{F})$ is finitely generated.*

Proof. Since K consists of finitely many simplices, the sequence $K_0 \subseteq K_1 \subseteq \cdots$ of inclusions eventually stabilises: that is, there exists $I \in \mathbb{N}$ such that $K_i = K_I$ for all $i \geq I$. Furthermore, since each K_i contains finitely many n-simplices, the $\mathbb{F}$-vector space $C_n(K_i; \mathbb{F})$ is finite-dimensional, hence so is $H_n(K_i; \mathbb{F})$; let $(e'_{i,1}, \ldots, e'_{i,m_i})$ be a basis for $H_n(K_i; \mathbb{F})$ over $\mathbb{F}$. Then $E' = \{e'_{i,j} \mid 0 \leq i \leq I, 1 \leq j \leq m_i\}$ is a finite generating set for $\bigoplus_{i=0}^{I} H_n(K_i; \mathbb{F})$ as an $\mathbb{F}$-vector space.

We claim that E' is also a generating set for $H_n(\{K_i\}, K; \mathbb{F})$ as an $\mathbb{F}[X]$-module. Indeed, since $K_i = K_{i+1}$ for $i \geq I$, the inclusion $f_i \colon K_i \to K_{i+1}$ induces an isomorphism in homology, implying that the map $H_n(K_I; \mathbb{F}) \to H_n(K_i; \mathbb{F})$ sending $y \mapsto X^{i-I} \cdot y$ is an isomorphism between $\mathbb{F}$-vector spaces as well. It follows that any element of $H_n(\{K_i\}, K; \mathbb{F})$ can be expressed in the form $\sum_{i=0}^{I-1} \sum_{j=1}^{m_i} \alpha_{i,j} \cdot e'_{i,j} + \sum_{j=1}^{m_I} P_j(X) \cdot e'_{I,j}$ for some $\alpha_{i,j} \in \mathbb{F}$ and $P_j(X) \in \mathbb{F}[X]$, proving our claim and the Proposition. □

Example. For each $i \in \mathbb{N}$, let $K_i = \{v_0, v_1, \ldots, v_{i+1}, v_0v_1, \ldots, v_{i-1}v_i\}$: that is, K_i is a "path of length i together with a point", e.g. $K_2 =$ 

; and let $f_i \colon K_i \to K_{i+1}$ be the inclusion. Let $\mathbb{F}$ be a field. One may then compute that $H_0(K_i; \mathbb{F}) = \mathbb{F}^2$ (in particular, $H_0(K_i; \mathbb{F})$ is finite-dimensional), and for a basis one can choose the images of e_{v_0} and $e_{v_{i+1}}$. With respect to this choice of basis, each of the maps $(f_i)_*$ induced in homology can be represented by the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, implying that the image of v_{i+1} in $H_0(K_i; \mathbb{F})$ is not of the form $\sum_{j=1}^{m} P_j(X) \cdot y_j$ for any $P_j(X) \in \mathbb{F}[X]$ and any $y_j \in \bigoplus_{i'=0}^{i-1} H_0(K_{i'}; \mathbb{F})$. This implies that the $\mathbb{F}[X]$-module $H_0(\{K_i\}, \{f_i\}; \mathbb{F})$ is not finitely generated.

3.2 Structure theorem for graded modules

Definition (cyclic graded module)**.** Let $\mathbb{F}$ be a field, let $b \in \mathbb{N}$, and let $d \in \mathbb{N} \cup \{\infty\}$ with $d > b$. A *cyclic graded* $\mathbb{F}[X]$*-module* with *birth* b and *death* d, denoted $C_{\mathbb{F}}(b, d)$, is the graded $\mathbb{F}[X]$-module $\bigoplus_{i=0}^{\infty} M_i$ such that

- given $i \in \mathbb{N}$, we have $M_i = \mathbb{F}$ if $b \leq i < d$ and $M_i = 0$ otherwise; and
- given $i, j \in \mathbb{N}$ with $i < j$, the map $M_i \to M_j, y \mapsto X^{j-i} \cdot y$ is an isomorphism if $b \leq i < j < d$ and the zero map otherwise.

Alternatively, one can think of $C_{\mathbb{F}}(b, d)$ as the quotient module $\frac{X^b \mathbb{F}[X]}{X^d \mathbb{F}[X]}$ (if $d < \infty$) or $X^b \mathbb{F}[X]$ (if $d = \infty$), where X^b is used to "shift the grading", and the quotient by the submodule $X^d \mathbb{F}[X]$ signifies "killing all grades above the d-th one".

Definition (graded isomorphism)**.** Let $\mathbb{F}$ be a field, and let $M = \bigoplus_{i=0}^{\infty} M_i$ and $N = \bigoplus_{i=0}^{\infty} N_i$ be two graded $\mathbb{F}[X]$-modules. An $\mathbb{F}$-linear map $f \colon M \to N$ is a *homomorphism* (of $\mathbb{F}[X]$-modules) if $f_{i+1}(X \cdot y) = X \cdot f_i(y)$ whenever $y \in M_i$, an *isomorphism* if it is in addition bijective, and a *graded isomorphism* if in addition $f(M_i) = N_i$ for all $i \in \mathbb{N}$. We say that M and N are *isomorphic* (resp. *graded isomorphic*), written $M \simeq N$ (resp. $M \cong N$), if there exists an isomorphism (resp. a graded isomorphism) $f \colon M \to N$.

Theorem (Structure Theorem)**.** *Let* $\mathbb{F}$ *be a field, and let* M *be a finitely generated graded* $\mathbb{F}[X]$*-module. Then there exist unique (up to permutations of indices)* $b_1, \ldots, b_s \in \mathbb{N}$ *and* $d_1, \ldots, d_s \in \mathbb{N} \cup \{\infty\}$*, with* $b_i < d_i$ *for each* i*, such that* $M \cong \bigoplus_{i=1}^{s} C_{\mathbb{F}}(b_i, d_i)$*.*

Sketch proof. The proof relies mostly on the fact that $R = \mathbb{F}[X]$ is a principal ideal domain (PID): this follows from the existence of a greatest common divisor for any two polynomials with coefficients in $\mathbb{F}$.

Since M is finitely generated, it has a finite generating set $\{x_1, \ldots, x_m\}$, implying that there exists a surjective R-module homomorphism $F\colon R^m \to M$ sending the basis elements to the x_i. In this case the kernel $K = \ker(F)$ is an R-module in its own right (we say K is a *submodule* of R^m), i.e. the map $R \times R^m \to R^m, (r, m) \mapsto r \cdot m$ restricts to a map $R \times K \to K$. As R is a PID, it can be shown that any submodule of a (finitely generated) free R-module is (finitely generated and) free, and therefore $K \simeq R^r$ for some $r \in \mathbb{N}$. The module M is then a *quotient R-module*, $M = R^m/K$.

Let A be an $m \times r$ matrix with entries in R representing the inclusion $R^r \simeq K \subseteq R^m$. It can be shown that A has a *Smith normal form*: that is, there exist invertible matrices P and Q and a diagonal $m \times r$ matrix D, all with entries in R, such that $A = PDQ$. (This is done by Gaussian elimination; we need to be slightly careful here, as R is only a PID and not a field, but in principle everything works the same way.) Since the matrices P and Q only correspond to changing the bases of R^m and R^r, respectively, they do not change the quotient R-module R^m/R^r up to isomorphism, implying that $M \simeq \bigoplus_{i=1}^s R/(d_i)$, where $d_1, \ldots, d_s$ are the diagonal entries of D that are not units in $R = \mathbb{F}[X]$ (i.e. the d_i are all diagonal entries of D apart from constant non-zero polynomials).

It now follows that $M \cong \bigoplus_{i=1}^s C_i$, where $R/(d_i) \simeq C_i \leq M$, and so it is enough to show that any graded $\mathbb{F}[X]$-module C that is isomorphic to $\mathbb{F}[X]/(P(X))$ for some non-constant polynomial $P(X) \in \mathbb{F}[X]$ is also graded isomorphic to a cyclic graded $\mathbb{F}[X]$-module. Writing $C = \bigoplus_{i=0}^\infty C_i$ for the grading, we must have $C_i \neq 0$ for some i (as $P(X)$ is non-constant), as well as $\dim_{\mathbb{F}}(C_i) \leq 1$ for all i and $C_i = 0$ whenever $C_j \neq 0$ and $C_k = 0$ for some $j < k < i$ (both properties follow since C is generated by a single element as an $\mathbb{F}[X]$-module). Writing $b := \min\{i \in \mathbb{N} \mid C_i \neq 0\}$ and $d := \min\{i > b \mid C_i = 0\}$ (or $d := \infty$ if $C_i \neq 0$ for all $i > b$), it then follows that $C \cong C_{\mathbb{F}}(b, d)$, as required.

Finally, to show uniqueness of the decomposition $M \cong \bigoplus_{i=1}^s C_{\mathbb{F}}(b_i, d_i)$, we use induction on s. In particular, if we write $M = \bigoplus_{i=0}^\infty M_i$ for the grading, and we set $b := \min\{i \in \mathbb{N} \mid M_i \neq 0\}$ and $d := \min\{i > b \mid X^{i-b} \cdot y = 0 \text{ for all } y \in M_b\}$ (or $d := \infty$ if the map $M_b \to M_i, y \mapsto X^{i-b} \cdot y$ is non-zero for all $i > b$), it can be shown that $(b, d) = (b_j, d_j)$ for some j and that $M/C_{\mathbb{F}}(b, d) \cong \bigoplus_{1 \leq i \leq s, i \neq j} C_{\mathbb{F}}(b_i, d_i)$. □

Example. Let $(\{K_i\}, K)$ be the filtration displayed in Figure 5, and let $\mathbb{F}$ be a field. As computed above, we have

$$H_1(K_1; \mathbb{F}) = \mathbb{F}(e_1'), \quad H_1(K_2; \mathbb{F}) = \mathbb{F}^2(e_2', e_3'), \quad H_1(K_3; \mathbb{F}) = \mathbb{F}(e_4'),$$

and $H_1(K_i; \mathbb{F}) = 0$ for $i \notin \{1, 2, 3\}$, and the $\mathbb{F}[X]$-module structure for $H_1(\{K_i\}, K; \mathbb{F})$ is defined by setting $X \cdot e_1' = e_2' + e_3'$, $X \cdot e_2' = 0$, $X \cdot e_3' = e_4'$ and $X \cdot e_4' = 0$. If we replace the $\mathbb{F}$-basis (e_2', e_3') for $H_1(K_2; \mathbb{F})$ with $(e_2', e_2' + e_3')$, it is then easy to see that we have $H_1(\{K_i\}, K; \mathbb{F}) \cong C_{\mathbb{F}}(1, 4) \oplus C_{\mathbb{F}}(2, 3)$, with the two cyclic graded $\mathbb{F}[X]$-modules generated by e_1' and e_2', respectively.

It can also be shown that $H_0(K_i; \mathbb{F}) = \mathbb{F}$ and that the map $(f_i)_*\colon H_0(K_i; \mathbb{F}) \to H_0(K_{i+1}; \mathbb{F})$ is an isomorphism for all $i \in \mathbb{N}$, implying that $H_0(\{K_i\}, K; \mathbb{F}) \cong C_{\mathbb{F}}(0, \infty)$.

3.3 The structure theorem in practice

As can be seen from the previous example, a decomposition of a graded module over $\mathbb{F}[X]$ into cyclic submodules can be thought of as a choice of particularly nice basis. In particular, the following is a consequence of the structure theorem.

Corollary. *Let $\mathbb{F}$ be a field, and let $M = \bigoplus_{i=0}^{\infty} M_i$ be a finitely generated graded $\mathbb{F}[X]$-module. Then, for each $i \geq 0$, there exists a basis $\mathcal{B}_i$ of M_i (as an $\mathbb{F}$-vector space) such that the following hold:*

- *for any $m \in \mathcal{B}_i$, we have $X \cdot m \in \mathcal{B}_{i+1} \sqcup \{0\}$;*
- *if $m, m' \in \mathcal{B}_i$ satisfy $X \cdot m = X \cdot m' \neq 0$, then $m = m'$.*

To see how a decomposition into cyclic submodules can be obtained from this, let $\widehat{\mathcal{B}}^{(0)} = \bigsqcup_{i=0}^{\infty} \mathcal{B}_i$, with $\mathcal{B}_i$ as above. Let $b = b_1 \geq 0$ be the smallest such that $\widehat{\mathcal{B}}^{(0)} \cap \mathcal{B}_b \neq \varnothing$, and let $m_1 \in \widehat{\mathcal{B}}^{(0)} \cap \mathcal{B}_b$. Then the $\mathbb{F}$-vector space V_1 generated by $\mathcal{C}_1 := \{X^j \cdot m_1 \mid j \geq 0\}$ is actually a cyclic $\mathbb{F}[X]$-module, and it follows from the Corollary that V_1 can be taken to be a direct summand of M. We may then set $\widehat{\mathcal{B}}^{(1)} := \widehat{\mathcal{B}}^{(0)} \setminus \mathcal{C}_1$ and repeat the procedure to obtain a cyclic submodule V_2, and so on; since M is finitely generated (as an $\mathbb{F}[X]$-module), we will eventually obtain $\widehat{\mathcal{B}}^{(r)} = \varnothing$, at which point we have a decomposition $M = \bigoplus_{i=1}^{r} V_r$.

In order to describe each V_i (generated, as an $\mathbb{F}$-vector space, by $\mathcal{C}_i := \{X^j \cdot m_i \mid j \geq 0\}$) as a cyclic module, note that we have $V_i \cong C_{\mathbb{F}}(b_i, d_i)$, where $b_i \geq 0$ is such that $m_i \in \mathcal{B}_{b_i}$, and d_i is the smallest integer $d > b_i$ such that $X^{d-b_i} \cdot m_i = 0$ (or $d_i = \infty$ if no such integer d exists). In particular, we have $d_i - b_i = \dim_{\mathbb{F}}(V_i)$.

The choice of the bases $\mathcal{B}_i$ as above may not always be unique. However, the pairs (b_i, d_i) obtained this way are determined uniquely by M: they do not depend on the choice of the $\mathcal{B}_i$.

Example. In the example above, set $v_1 = e_1'$, $v_2 = e_2' + e_3'$, $v_2' = e_2'$ and $v_3 = e_4'$. The bases $\mathcal{B}_1 = \{v_1\}$, $\mathcal{B}_2 = \{v_2, v_2'\}$, $\mathcal{B}_3 = \{v_3\}$ and $\mathcal{B}_i = \varnothing$ for $i \notin \{1,2,3\}$ satisfy the conclusion of the Corollary: indeed, we have $X \cdot v_1 = v_2$, $X \cdot v_2 = v_3$, $X \cdot v_2' = 0$ and $X \cdot v_3 = 0$. Applying the above algorithm then results in $H_1(\{K_i\}, K; \mathbb{F}) \cong V_1 \oplus V_2$, where $V_1 = \mathrm{span}_{\mathbb{F}}\{v_1, v_2, v_3\} \cong C_{\mathbb{F}}(1,4)$ and $V_2 = \mathrm{span}_{\mathbb{F}}\{v_2'\} \cong C_{\mathbb{F}}(2,3)$.

3.4 Persistence diagrams and barcodes

A multiset in a set X is a generalisation of a subset of X in which points are allowed to appear "with multiplicity", and a persistence diagram is a multiset of points (b, d) with $0 \leq b < d \leq \infty$, defined as follows. Here we write $[0, \infty]$ for the set $[0, \infty) \cup \{\infty\}$.

Definition (persistence diagram)**.**

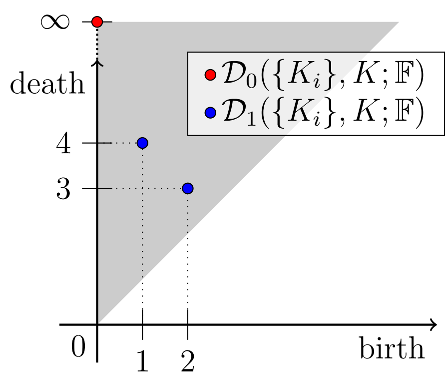
- Let X be a set. A (*finite*) *multiset* in X is a collection of points $\{\{x_1, \ldots, x_s\}\}$ in X, where two multisets $\{\{x_1, \ldots, x_s\}\}$ and $\{\{y_1, \ldots, y_t\}\}$ are considered to be the same if and only if $s = t$ and $(x_1, \ldots, x_s)$ is a permutation of $(y_1, \ldots, y_t)$.

- Given $x \in X$ and a multiset σ in X, the *multiplicity* of x in σ is the number of times x appears in σ.
- A *persistence diagram* is a finite multiset in $\Delta := \{(b, d) \in [0, \infty]^2 \mid b < d\}$.

Now suppose that $\mathbb{F}$ is a field and M is a finitely generated graded $\mathbb{F}[X]$-module. Then M uniquely decomposes as $M \cong \bigoplus_{i=1}^{s} C_{\mathbb{F}}(b_i, d_i)$ by the Structure Theorem. The *persistence diagram* of M, denoted $\mathcal{D}(M)$, is the persistence diagram $\{\{(b_1, d_1), \ldots, (b_s, d_s)\}\}$. More generally, given $\varepsilon > 0$, the *step-ε persistence diagram* of M, denoted $\mathcal{D}^{\varepsilon}(M)$, is the persistence diagram $\{\{(\varepsilon b_1, \varepsilon d_1), \ldots, (\varepsilon b_s, \varepsilon d_s)\}\}$, so that $\mathcal{D}^1(M) = \mathcal{D}(M)$.

We will be particularly interested in the situation when $M = H_n(\{K_i\}, K; \mathbb{F})$ for some filtration $(\{K_i\}, K)$ of simplicial complexes and some $n \in \mathbb{N}$. In that case, we will write $\mathcal{D}_n(\{K_i\}, K; \mathbb{F})$ and $\mathcal{D}_n^{\varepsilon}(\{K_i\}, K; \mathbb{F})$ for $\mathcal{D}(M)$ and $\mathcal{D}^{\varepsilon}(M)$, respectively.

Example. Let $(\{K_i\}, K)$ be the filtration displayed in Figure 5, and let $\mathbb{F}$ be a field. As computed above, we have $H_0(\{K_i\}, K; \mathbb{F}) \cong C_{\mathbb{F}}(0, \infty)$ and $H_1(\{K_i\}, K; \mathbb{F}) \cong C_{\mathbb{F}}(1, 4) \oplus C_{\mathbb{F}}(2, 3)$, giving rise to persistence diagrams $\{\{(0, \infty)\}\}$ and $\{\{(1, 4), (2, 3)\}\}$, respectively. Such diagrams are drawn on the right, in red and blue, respectively; it is common to draw several persistence diagrams $\mathcal{D}_n(\{K_i\}, K; \mathbb{F})$ (for different values of n) on the same graph using different colours, as is done here. The shading indicates the region Δ where points of any given persistence diagram might lie.

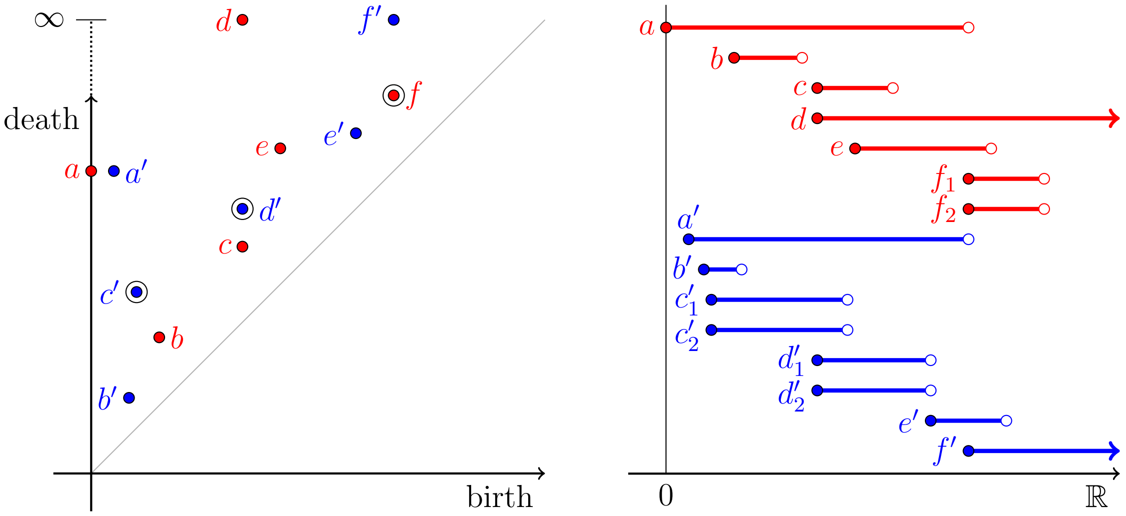


We may want to use step-ε persistence diagrams (for values of $\varepsilon > 0$ other than $\varepsilon = 1$) for Morse, Čech or Rips filtrations with step ε. In fact, it is not hard to see how each of Morse, Čech and Rips filtration constructions could be modified to yield a collection $\{K_x \mid x \in [0, \infty)\}$ of subcomplexes of a simplicial complex K satisfying $K_x \subseteq K_y$ for $x \leq y$, so that the corresponding filtration with step ε is precisely $\mathcal{K}_{\varepsilon} := (\{K_{i\varepsilon} \mid i \in \mathbb{N}\}, K)$. It can then be shown that the persistence diagrams $\mathcal{D}_n^{\varepsilon}(\mathcal{K}_{\varepsilon}; \mathbb{F})$ converge (in a certain sense) to another persistence diagram as $\varepsilon \to 0$. This "limit" persistence diagram is what we usually consider when we are talking about persistent homology of Morse, Rips or Čech filtrations.

An alternative representation of a persistence diagram is a *barcode*, where we replace each point (b, d) in a persistence diagram with the half-open interval $[b, d) \subset \mathbb{R}$, and plot these intervals on top of each other.

3.5 Bottleneck distance

One of the most useful features of persistent diagrams is their universality: regardless of what data we have and what method we use for constructing a filtration of simplicial complexes, the end result is the same—a multiset of points in $[0, \infty]^2$. We can therefore use



(a) Persistence diagrams $\mathcal{D}_0$ (red) and $\mathcal{D}_1$ (blue). (b) The corresponding barcode.

Figure 9: The construction of a barcode from a pair of persistence diagrams. Points denoted by ⦿ and ⦿ indicate the points of $\mathcal{D}_0$ and $\mathcal{D}_1$ (respectively) of multiplicity 2; in general, points of higher multiplicity are also possible in a persistence diagram.

persistence diagrams to compare datasets that would be difficult to compare otherwise, such as pictures of different resolutions (using Morse filtrations) or subspaces of different metric spaces (using Čech or Rips filtrations). In order to do that, we need to introduce a measure of dissimilarity, or distance, between two persistence diagrams. We first need a slight modification of the definition of persistence diagrams.

Definition (persistence diagram, version 2)**.** Let $\Delta_= = \{(b,d) \in [0,\infty]^2 \mid b \leq d\}$.

- We say a point $(b,d) \in \Delta_=$ is *diagonal* if $b = d$, and *off-diagonal* if $b < d$.
- We say two multisets σ and τ in $\Delta_=$ are *equivalent*, written $\sigma \sim \tau$, if we can write $\sigma \sqcup \sigma' = \tau \sqcup \tau'$, where σ' and τ' are finite multisets consisting of diagonal points. This can be seen to be an equivalence relation; we write $[\sigma]$ for the equivalence class of σ.
- A *persistence diagram* is an equivalence class $[\sigma]$ of some multiset σ in $\Delta_=$.

This definition is different from the previous one in that it allows diagonal points, but assumes some different multisets represent the same persistence diagram. However, the definitions are easily seen to be equivalent: if $[\sigma]$ is a persistence diagram in the "new" sense then it represents a unique persistence diagram σ' in the "old" sense, where σ' consists precisely of the off-diagonal points in σ.

Definition (bottleneck distance)**.** Let σ and τ be finite multisets in $\Delta_=$.

- Given a constant $\varepsilon \geq 0$, an *ε-matching* between σ and τ is a bijection $\phi\colon \sigma \to \tau$ such that $|b - b'| \leq \varepsilon$ and $|d - d'| \leq \varepsilon$ for all $(b,d) \in \sigma$, where we write (b',d') for $\phi(b,d)$. (By convention, we assume that $|\infty - \infty| = 0$ here.)

- The *bottleneck distance* between the persistence diagrams $[\sigma]$ and $[\tau]$, which we denote by $d_\infty([\sigma],[\tau])$, is the infimum of all $\varepsilon \geq 0$ such that there exist multisets $\sigma' \sim \sigma$ and $\tau' \sim \tau$ together with an ε-matching between σ' and τ'.

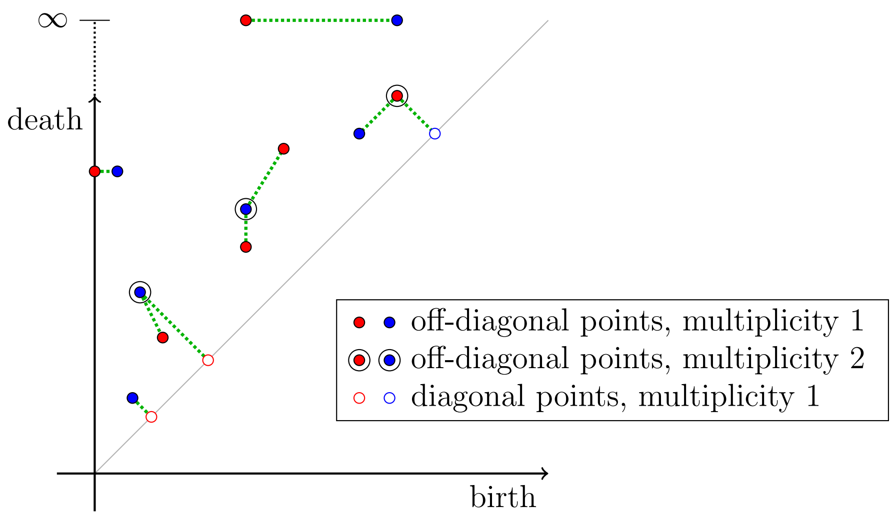


Figure 10: An illustration of bottleneck distance between two diagrams σ and τ (represented by red and blue points, respectively). The green dotted lines define an ε-matching between diagrams $\sigma' \sim \sigma$ and $\tau' \sim \tau$, for any constant ε satisfying $\varepsilon \geq |b-b'|$ and $\varepsilon \geq |d-d'|$ for any pair (b,d), (b',d') of points matched.

Note that multisets $\sigma' \sim \sigma$ and $\tau' \sim \tau$ that have the same number of elements always exist: we could simply add a suitable number of copies of $(0,0)$ to σ or to τ, say. Therefore, we have $d_\infty([\sigma],[\tau]) = \infty$ if and only if σ and τ contain a different number of points (x,∞) for elements $x \in [0,\infty)$. Moreover, if $d_\infty([\sigma],[\tau]) < \infty$ then "infimum" can be replaced by "minimum" in the definition above—that is, given two persistence diagrams $[\sigma]$ and $[\tau]$, it can be shown that there exists an ε-matching between σ' and τ', where $\sigma' \sim \sigma$, $\tau' \sim \tau$ and $\varepsilon = d_\infty([\sigma],[\tau])$.

Proposition. *The function d_∞ defines an extended metric on the set of all persistence diagrams. (Here,* extended *means that we allow d_∞ to take the value ∞.)*

Proof. It is clear that d_∞ is reflexive ($d_\infty([\sigma],[\sigma]) = 0$) and symmetric ($d_\infty([\sigma],[\tau]) = d_\infty([\tau],[\sigma])$). We therefore only need to show that d_∞ satisfies the triangle inequality ($d_\infty([\sigma],[\tau]) + d_\infty([\tau],[\rho]) \leq d_\infty([\sigma],[\rho])$).

Let σ, τ and ρ be multisets in $\Delta_=$, and suppose that $d_\infty([\sigma],[\tau]) < \alpha$ and $d_\infty([\tau],[\rho]) < \beta$ for some $\alpha, \beta \in (0,\infty)$. It is then enough to show that $d_\infty([\sigma],[\rho]) \leq \alpha + \beta$. Now there exist $\sigma' \sim \sigma$, $\tau' \sim \tau$ and an α-matching $\phi\colon \sigma' \to \tau'$, as well as $\tau'' \sim \tau$, $\rho'' \sim \rho$ and a β-matching $\psi\colon \tau'' \to \rho''$. By the definition of the equivalence relation $\sim$, we can write $\sigma \sqcup \hat\sigma = \sigma' \sqcup \hat\sigma'$, $\tau \sqcup \hat\tau = \tau' \sqcup \hat\tau'$, $\tau \sqcup \hat\tau_0 = \tau'' \sqcup \hat\tau''$ and $\rho \sqcup \hat\rho = \rho'' \sqcup \hat\rho''$, where $\hat\sigma$, $\hat\sigma'$, $\hat\tau$, $\hat\tau'$, $\hat\tau_0$, $\hat\tau''$, $\hat\rho$ and $\hat\rho''$ are finite multisets consisting of diagonal points. Note that we have

$$\tau' \sqcup \hat\tau' \sqcup \hat\tau_0 = \tau \sqcup \hat\tau \sqcup \hat\tau_0 = \tau'' \sqcup \hat\tau'' \sqcup \hat\tau;$$

we can therefore consider the composite bijection $\hat{\psi} \circ \hat{\phi}$, where $\hat{\phi}\colon \sigma' \sqcup \hat{\tau}' \sqcup \hat{\tau}_0 \to \tau' \sqcup \hat{\tau}' \sqcup \hat{\tau}_0$ and $\hat{\psi}\colon \tau'' \sqcup \hat{\tau}'' \sqcup \hat{\tau} \to \rho'' \sqcup \hat{\tau}'' \sqcup \hat{\tau}$ are bijections defined by

$$\hat{\phi}(b,d) = \begin{cases} \phi(b,d) & \text{if } (b,d) \in \sigma', \\ (b,d) & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \hat{\psi}(b,d) = \begin{cases} \psi(b,d) & \text{if } (b,d) \in \tau'', \\ (b,d) & \text{otherwise.} \end{cases}$$

It is easy to see that $\hat{\psi} \circ \hat{\phi}$ is an $(\alpha+\beta)$-matching between $\sigma' \sqcup \hat{\tau}' \sqcup \hat{\tau}_0$ and $\rho'' \sqcup \hat{\tau}'' \sqcup \hat{\tau}$, and that we have $\sigma \sim \sigma' \sqcup \hat{\tau}' \sqcup \hat{\tau}_0$ and $\rho \sim \rho'' \sqcup \hat{\tau}'' \sqcup \hat{\tau}$, as required. □

The following result motivates our definition of bottleneck distance—in particular, it motivates allowing matching points to the diagonal $\{(x,x) \mid x \in [0,\infty]\}$. Points close to the diagonal are usually a consequence of noise and may, to a certain extent, be ignored (even though there might be lots of such points in a persistence diagram coming from real-world data.

Theorem (Stability of persistence)**.** *Let K be a simplicial complex, and let $f, g\colon K \to [0,\infty)$ be combinatorial Morse functions. Pick $n \geq 0$, and let $\mathbb{F}$ be a field. Let $\mathcal{D}_f$ be the persistence diagram in dimension n over $\mathbb{F}$ associated to f, i.e. the limit (as $\varepsilon \to 0$) of diagrams $\mathcal{D}_n^\varepsilon(\{K_{i\varepsilon} \mid i \in \mathbb{N}\}, K; \mathbb{F})$, where $K_x = \{\sigma \in K \mid f(\sigma) \leq x\}$; define $\mathcal{D}_g$ similarly. Then*

$$d_\infty(\mathcal{D}_f, \mathcal{D}_g) \leq \|f-g\|_\infty := \max_{\sigma \in K} |f(\sigma) - g(\sigma)|.$$

3.6 Wasserstein distances

It turns out that bottleneck distance is just a special case of a whole two-parameter family of metrics, called Wasserstein distances and defined as follows.

Definition ((p,q)-Wasserstein distance)**.** Let $p, q \in [1,\infty]$.

- Given two points $(b,d), (b',d') \in \Delta_=$, the *$q$-distance* between (b,d) and (b',d') is defined as
$$d'_q((b,d),(b',d')) := \left(|b-b'|^q + |d-d'|^q\right)^{1/q}$$
if $q < \infty$, and
$$d'_\infty((b,d),(b',d')) := \max\{|b-b'|, |d-d'|\}$$
for $q = \infty$.

- Given two multisets σ', τ' in $\Delta_=$, we define the *(p,q)-Wasserstein distance* between the persistence diagrams $[\sigma]$ and $[\tau]$ as
$$d_{p,q}([\sigma],[\tau]) := \inf_{\phi\colon \sigma' \to \tau'} \left(\sum_{x \in \sigma'} d'_q(x, \phi(x))^p \right)^{1/p} \qquad (*)$$
if $p < \infty$, and
$$d_{\infty,q}([\sigma],[\tau]) := \inf_{\phi\colon \sigma' \to \tau'} \max_{x \in \sigma'} d'_q(x, \phi(x)) \qquad (**)$$
for $p = \infty$. Here, the infimum is taken over all bijections $\phi\colon \sigma' \to \tau'$ with $\sigma' \sim \sigma$ and $\tau' \sim \tau$.

Given this definition, the bottleneck distance is then nothing else than the (∞, ∞)-Wasserstein distance. This explains the notation d_∞ for the bottleneck distance (although we would write $d_{\infty,\infty}$ in this case).

Now we turn our discussion towards computing Wasserstein (and, in particular, bottleneck) distances in practice. In theory, this seems to be a hard problem, as one may *a priori* need to consider infinitely many bijections $\phi\colon \sigma' \to \tau'$ (one may consider forming σ' by adding (x, x) to σ for any $x \in [0, \infty]$, for instance). However, it turns out that considering that many possibilities is not necessary, as we only need to check the (finitely many) bijections given by the following definition.

Definition (potentially optimal bijection)**.** Let σ and τ be persistence diagrams. A bijection $\phi\colon \sigma' \to \tau'$ between two persistence diagrams is said to be *potentially optimal* for (σ, τ) if the following conditions hold:

- $\sigma \sim \sigma'$ and $\tau \sim \tau'$;
- for any $x = (b, d) \in \sigma'$ with $\phi(x) = (b', d') \in \tau'$, either $b < d$ or $b' < d'$ (or both);
- if we have $\phi(x) = (c', c') \in \tau'$ for some $c' \in [0, \infty]$, where $x = (b, d) \in \sigma'$, then $c' = \frac{b+d}{2}$; and
- if we have $x = (c, c) \in \sigma'$ for some $c \in [0, \infty]$, and $\phi(x) = (b', d') \in \tau'$, then $c = \frac{b'+d'}{2}$.

In words, the definition says that a bijection is potentially optimal if it does not match diagonal points to other diagonal points, and if whenever a point x is matched to a diagonal point, it is matched to the diagonal point that is closest to x. It is straightforward that only potentially optimal bijections need to be considered in the computations of Wasserstein distances: indeed, a non-(potentially optimal) bijection $\phi\colon \sigma' \to \tau'$, can be replaced by a "better" bijection $\phi'\colon \sigma'' \to \tau''$, by either removing pairs of matched diagonal points, or replacing a diagonal point x matched to a non-diagonal point y with another diagonal point x' satisfying $d'_q(x', y) \leq d'_q(x, y)$. In particular, in equations $(*)$ and $(**)$, the infimum only needs to be taken over potentially optimal bijections for (σ, τ), and since there are only finitely many of those, we may furthermore replace $\inf\limits_{\phi\colon \sigma' \to \tau'}$ with $\min\limits_{\phi\colon \sigma' \to \tau'}$.

Example. Consider the following persistence diagrams: $\sigma = \{\{(0, 6), (0, 6)\}\}$ and $\tau = \{\{(0, 2), (5, 9)\}\}$. Then a potentially optimal bijection can either match both (off-diagonal) points of σ to the diagonal point $(3, 3)$, or one point of σ to $(3, 3)$ and the other one to an (off-diagonal) point of τ, or both points of σ to points of τ. This results in the following list of potentially optimal bijections for (σ, τ); see Figure 11:

$$
\begin{array}{llll}
\phi_1\colon (0,6) \mapsto (3,3), & \phi_2\colon (0,6) \mapsto (0,2), & \phi_3\colon (0,6) \mapsto (5,9), & \phi_4\colon (0,6) \mapsto (0,2). \\
\quad\;\; (0,6) \mapsto (3,3) & \quad\;\; (0,6) \mapsto (3,3) & \quad\;\; (0,6) \mapsto (3,3) & \quad\;\; (0,6) \mapsto (5,9) \\
\quad\;\; (1,1) \mapsto (0,2) & \quad\;\; (7,7) \mapsto (5,9) & \quad\;\; (1,1) \mapsto (0,2) & \\
\quad\;\; (7,7) \mapsto (5,9) & & &
\end{array}
$$

Now given a potentially optimal bijection $\phi\colon \sigma' \to \tau'$ and constants $p, q \in [1, \infty]$, one may define $c_{p,q}(\phi) = \left(\sum_{x \in \sigma'} d'_q(x, \phi(x))^p\right)^{1/p}$ if $p < \infty$ and $c_{\infty,q}(\phi) = \max_{x \in \sigma'} d'_q(x, \phi(x))$. We then have $d_{p,q}([\sigma],[\tau]) = \min\{c_{p,q}(\phi_i) \mid 1 \leq i \leq 4\}$. In particular, we may compute some Wasserstein distances $d_{p,q}([\sigma],[\tau])$, as follows.

(p,q)	$c_{p,q}(\phi_1)$	$c_{p,q}(\phi_2)$	$c_{p,q}(\phi_3)$	$c_{p,q}(\phi_4)$	$d_{p,q}([\sigma],[\tau])$
(∞,∞)	3	4	5	5	3
$(\infty,1)$	6	6	8	8	6
$(1,\infty)$	9	9	9	9	9
$(1,1)$	18	14	16	12	12
$(2,2)$	$\sqrt{46}$	$\sqrt{42}$	$\sqrt{54}$	$\sqrt{50}$	$\sqrt{42}$

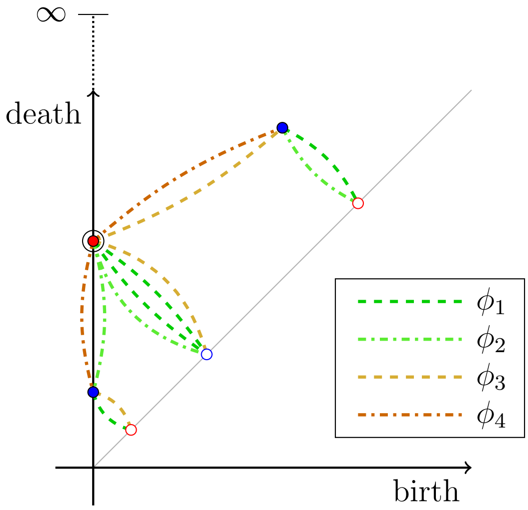


Figure 11: Potentially optimal bijections for the pair (σ,τ), where $\sigma = \{\{(0,6),(0,6)\}\}$ and $\tau = \{\{(0,2),(5,9)\}\}$.