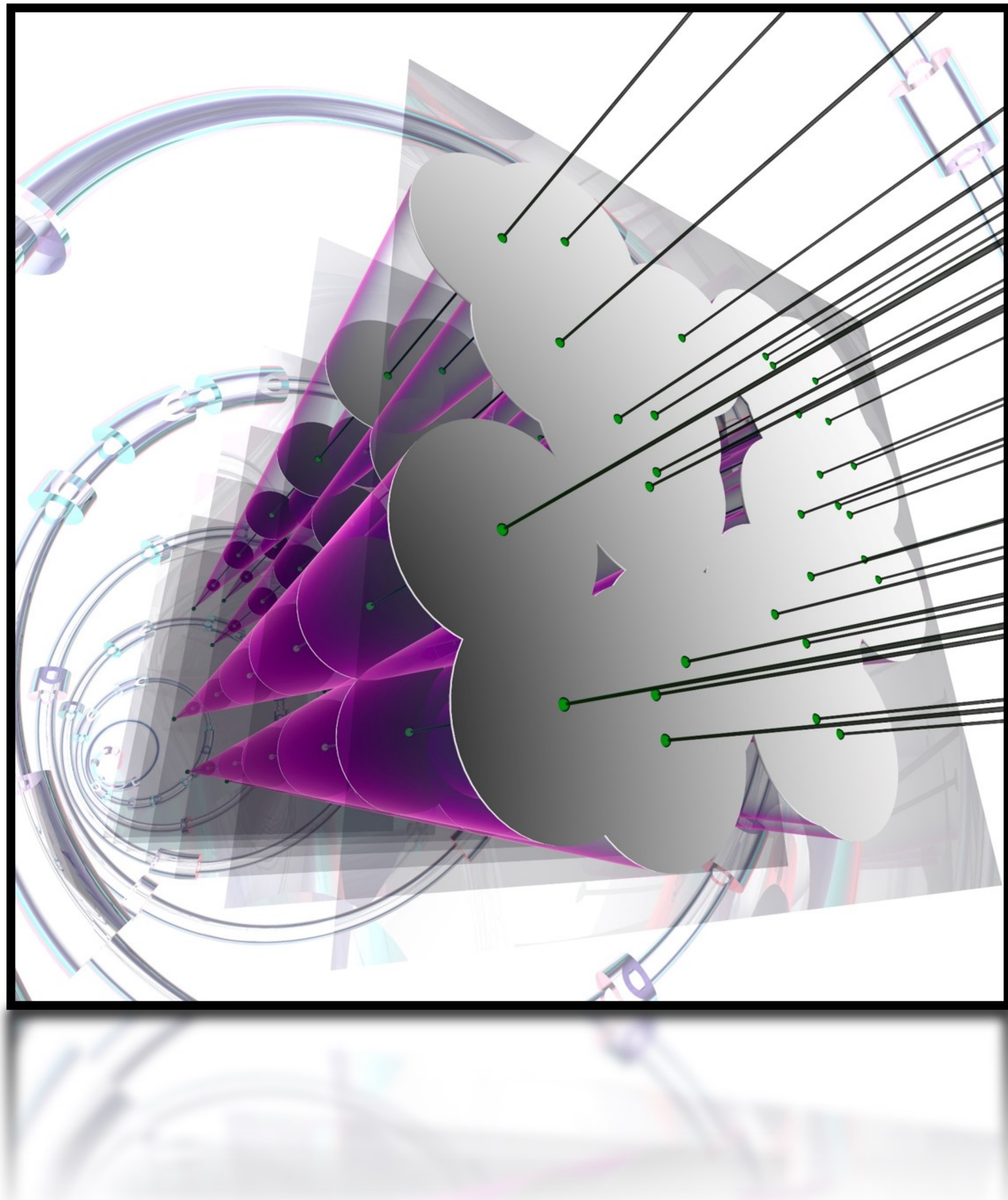
6. Persistence

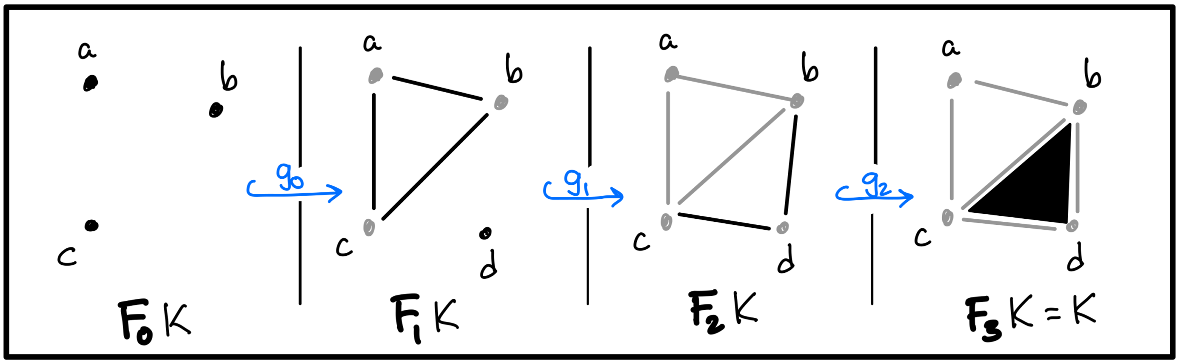


6.1 Persistent Homology

Consider a filtration $\boldsymbol{F}_\bullet K$ of a simplicial complex K (as in Definition 1.6):

$$\boldsymbol{F}_0 K \subset \boldsymbol{F}_1 K \subset \cdots \subset \boldsymbol{F}_{n-1} K \subset \boldsymbol{F}_n K = K,$$

and denote the inclusion simplicial maps by $g_i : \boldsymbol{F}_i K \hookrightarrow \boldsymbol{F}_{i+1} K$. Here is one such filtration:



There are induced linear maps on homology $\mathbf{H}_k g_i : \mathbf{H}_k(\boldsymbol{F}_i K) \to \mathbf{H}_k(\boldsymbol{F}_{i+1} K)$ in every dimension $k \geq 0$ (see Sections 2 and 3 of Chapter 4). For a fixed k, these linear maps fit together into a sequence of vector spaces:

$$\mathbf{H}_k(\boldsymbol{F}_0 K) \xrightarrow{\mathbf{H}_k g_0} \mathbf{H}_k(\boldsymbol{F}_1 K) \xrightarrow{\mathbf{H}_k g_1} \cdots \xrightarrow{\mathbf{H}_k g_{n-2}} \mathbf{H}_k(\boldsymbol{F}_{n-1} K) \xrightarrow{\mathbf{H}_k g_{n-1}} \mathbf{H}_k(\boldsymbol{F}_n K).$$

There are several other induced maps on homology hiding in plain sight — for instance, we have said nothing about the inclusion $g_1 \circ g_0 : \boldsymbol{F}_0 K \hookrightarrow \boldsymbol{F}_2 K$. Fortunately for us, homology is functorial (see Proposition 4.8); so the missing map $\mathbf{H}_k(g_1 \circ g_0)$ is easily reconstructed by composing the available maps $\mathbf{H}_k g_1$ and $\mathbf{H}_k g_0$.

More generally, for any pair $i < j$ of filtration indices in $\{0, \ldots, n\}$, the map induced on homology by the inclusion $g_{i\to j} : \boldsymbol{F}_i K \hookrightarrow \boldsymbol{F}_j K$ is the composite $\mathbf{H}_k(\boldsymbol{F}_i K) \to \mathbf{H}_k(\boldsymbol{F}_j K)$ in our diagram of vector spaces, i.e.,

$$\mathbf{H}_k g_{i\to j} = \mathbf{H}_k g_{j-1} \circ \mathbf{H}_k g_{j-1} \circ \cdots \circ \mathbf{H}_k g_{i+1} \circ \mathbf{H}_k g_i.$$

Such maps $\mathbf{H}_k g_{i\to j}$ contain crucial information which allows us to coherently connect the k-th homology groups of *all* the subcomplexes which appear in the filtration $\boldsymbol{F}_\bullet$ of K. The key point is that in order to extract this information, we must study sequences of vector spaces; thus, we are inexorably led to the following definition.

Definition 6.1. An $\mathbb{N}$-indexed **persistence module** over $\mathbb{F}$ is a sequence $(V_\bullet, a_\bullet)$ of $\mathbb{F}$-vector spaces V_k and linear maps a_k defined for $k \geq 0$ which fit into a diagram

$$V_0 \xrightarrow{a_0} V_1 \xrightarrow{a_1} V_2 \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} V_k \xrightarrow{a_k} V_{k+1} \xrightarrow{a_{k+1}} \cdots$$

The maps $a_\bullet$ are not required to satisfy $a_k \circ a_{k-1} = 0$, so persistence modules need not be cochain complexes (compare Definition 5.1); conversely, every cochain complex is automatically a persistence module. In any event, for every pair $i \leq j$ in $\mathbb{N}$ we will write the composite map $a_{j-1} \circ a_{j-2} \circ \cdots \circ a_i$ via the shorthand $a_{i\to j} : V_i \to V_j$, with the implicit understanding that $a_{i\to i}$ is just the identity map on V_i.

Remark 6.2. We say that a persistence module $(V_\bullet, a_\bullet)$ is of **finite type** if $\dim V_i < \infty$ for all $i \geq 0$ and if the maps $a_i : V_i \to V_{i+1}$ are isomorphisms for all $i \gg 0$. Both these conditions are satisfied by persistence modules obtained from homology groups of filtered simplicial complexes.

We now turn to the objects of interest.

DEFINITION 6.3. For each pair $i \leq j$ of integers, the associated **persistent homology group** of a persistence module $(V_\bullet, a_\bullet)$ is the subspace of V_j given by

$$\mathbf{PH}_{i\to j}(V_\bullet, a_\bullet) = \mathrm{img}(a_{i\to j}).$$

It is not too difficult to check that $\mathbf{PH}_{i\to j}(V_\bullet, a_\bullet)$ is a subset of $\mathbf{PH}_{i'\to j}(V_\bullet, a_\bullet)$ whenever $i' \geq i$. We say that a vector v in V_i is **born** at filtration index i if v does not lie in $\mathrm{img}\, a_{i-1}$; similarly, v is said to **die** at filtration index $j \geq i$ whenever j is the *smallest* number satisfying $a_{i\to j}(v) = 0$; by convention, the death index of v equals $+\infty$ no such j exists, i.e., if $a_{i\to j}(v)$ is nonzero for all $j \geq i$. The **persistence** of v is defined to be death minus birth, i.e., $(j - i)$.

REMARK 6.4. In the special case where our persistence module arises from taking the k-th homology groups of a filtered simplicial complex as described above, we will denote the persistent homology groups as $\mathbf{PH}_k g_{i\to j}(\boldsymbol{F}_\bullet K)$ for all $i \leq j$. The group $\mathbf{PH}_k g_{i\to j}(\boldsymbol{F}_\bullet K)$ consists of precisely those homology classes in $\mathbf{H}_k(\boldsymbol{F}_i K)$ which continue to generate nontrivial homology in the larger complex $\boldsymbol{F}_j K$ — geometrically, these are precisely those (equivalence classes of) k-cycles in $\boldsymbol{F}_i K$ which do not become k-boundaries in $\boldsymbol{F}_j K$. Writing ∂_k^i for the k-th boundary operator of each simplicial complex $\boldsymbol{F}_i K$, we have

$$\mathbf{PH}_k g_{i\to j}(\boldsymbol{F}_\bullet K) = \mathbf{H}_k g_{i\to j}(\ker \partial_k^i) \big/ [\mathbf{H}_k g_{i\to j}(\ker \partial_k^i) \cap \mathrm{img}\, \partial_{k+1}^j].$$

And in particular, $\mathbf{PH}_k g_{i\to i}(\boldsymbol{F}_\bullet K)$ is just the k-th homology group of $\boldsymbol{F}_i K$.

The study of persistence modules is greatly facilitated by two miracles — an inherently algebraic **structure theorem** and a viscerally geometric **stability theorem**. The first of these allows us to represent every persistence module using the combinatorial data called its *barcode*. And the stability theorem asserts that the assignment of barcodes to modules is an isometry under certain natural metrics. We will describe the structure theorem in the next section

6.2 BARCODES

The quest to understand persistent homology groups begins, like many good quests, with the establishment of a categorical framework.

DEFINITION 6.5. A **morphism** between persistence modules $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ is a family of linear maps $\phi_k : V_k \to W_k$ which satisfy

$$b_i \circ \phi_i = \phi_{i+1} \circ a_i$$

for every $i \geq 0$

This definition amounts to requiring the commutativity of all squares in the following diagram of vector spaces:

$$\begin{tikzcd}
V_0 \arrow[r, "a_0"] \arrow[d, "\phi_0"'] & V_1 \arrow[r, "a_1"] \arrow[d, "\phi_1"'] & \cdots \arrow[r, "a_{k-1}"] & V_k \arrow[r, "a_k"] \arrow[d, "\phi_k"'] & V_{k+1} \arrow[r, "a_{k+1}"] \arrow[d, "\phi_{k+1}"'] & \cdots \\
W_0 \arrow[r, "b_0"'] & W_1 \arrow[r, "b_1"'] & \cdots \arrow[r, "b_{k-1}"'] & W_k \arrow[r, "b_k"'] & W_{k+1} \arrow[r, "b_{k+1}"'] & \cdots
\end{tikzcd}$$

The pair (persistence modules, their morphisms) forms a category in the sense of Definition 4.1. We call $\phi_\bullet : (V_\bullet, a_\bullet) \to (W_\bullet, b_\bullet)$ an *isomorphism* if every ϕ_i is an invertible linear map of vector spaces in the usual sense. If such an isomorphism exists, we write $(V_\bullet, a_\bullet) \simeq (W_\bullet, b_\bullet)$.

DEFINITION 6.6. The **direct sum** of two persistence modules $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ is a new persistence module $(V_\bullet \oplus W_\bullet, a_\bullet \oplus b_\bullet)$ defined as follows: its k-th vector space is the direct sum $V_i \oplus W_i$, while the linear map $a_i \oplus b_i$ has matrix representation $\left[\begin{smallmatrix} a_i & 0 \\ 0 & b_i \end{smallmatrix}\right]$.

Persistent homology groups of direct sums are direct sums of persistent homology groups (see Exercise 6.1 of this Chapter for a precise statement.) We say that a persistence module $(I_\bullet, c_\bullet)$ is **indecomposable** if it does not admit any interesting direct sum decompositions — in other words, anytime we have an isomorphism

$$(I_\bullet, c_\bullet) \simeq (V_\bullet, a_\bullet) \oplus (W_\bullet, b_\bullet),$$

of persistence modules, one of the factors on the right side will be isomorphic to $(I_\bullet, c_\bullet)$, while the other one will be zero everywhere. The following result highlights a particularly important class of indecomposable persistent modules.

PROPOSITION 6.7. *Let $(I_\bullet, c_\bullet)$ be a nonzero $\mathbb{N}$-indexed persistence module over a field $\mathbb{F}$. Assume that there exist indices $i \leq j$ with i in $\mathbb{N}$ and j in $\mathbb{N} \cup \{\infty\}$ so that*

$$\dim I_p = \begin{cases} 1 & i \leq p \leq j \\ 0 & \textit{otherwise} \end{cases}, \quad \textit{and} \quad \operatorname{rank}\left(c_p : I_p \to I_{p+1}\right) = \begin{cases} 1 & i \leq p < j \\ 0 & \textit{otherwise} \end{cases}.$$

Then, $(I_\bullet, c_\bullet)$ is indecomposable.

PROOF. Consider any direct sum decomposition $(I_\bullet, c_\bullet) \simeq (V_\bullet, a_\bullet) \oplus (W_\bullet, b_\bullet)$. For each p in $\{i, i+1, \ldots, j\}$ we have $\dim V_p + \dim W_p = \dim I_p = 1$; let's assume without loss of generality that $\dim V_i = 1$ and $\dim W_i = 0$. This forces the map b_i to be zero, and by Definition 6.5 we now have a commutative diagram which looks like:

$$\begin{tikzcd} \mathbb{F} \arrow[r, "\simeq"] \arrow[d, "\simeq"'] & \mathbb{F} \arrow[d, "\simeq"] \\ \mathbb{F} \oplus 0 \arrow[r, "\left[\begin{smallmatrix} a_i & 0 \\ 0 & 0 \end{smallmatrix}\right]"'] & V_{i+1} \oplus W_{i+1} \end{tikzcd}$$

with all arrows labelled $\simeq$ being vector space isomorphisms. It follows that a_i has rank one, $\dim V_{i+1} = 1$, and $\dim W_{i+1} = 0$. Continuing onwards by induction on i, we see that $(V_\bullet, a_\bullet)$ is isomorphic to $(I_\bullet, c_\bullet)$ while $(W_\bullet, b_\bullet)$ is trivial; thus, $(I_\bullet, c_\bullet)$ is indecomposable as desired. □

Up to isomorphism, every indecomposable module of the form described in the proposition above is completely characterized by knowledge of the pair of integers $i \leq j$ (allowing for the fact that j might equal ∞).

DEFINITION 6.8. For each pair $0 \leq i \leq j \leq \infty$ (with $i \neq \infty$), the $\mathbb{N}$-indexed **interval module** $(I_\bullet^{i,j}, c_\bullet^{i,j})$ over $\mathbb{F}$ is given by

$$I_p^{i,j} = \begin{cases} \mathbb{F} & i \leq p \leq j \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \quad c_p^{i,j} = \begin{cases} \mathrm{id}_{\mathbb{F}} & i \leq p < j \\ 0 & \text{otherwise} \end{cases}.$$

(Here $\mathrm{id}_{\mathbb{F}}$ denotes the identity map $\mathbb{F} \to \mathbb{F}$).

The first miracle of persistent homology is the following result, which allows us to uniquely express *any* $\mathbb{N}$-indexed persistence module of finite type as a direct sum of finitely many interval modules. Please do not panic (yet) if various terms in the proof appear intimidating — clarifying remarks and concrete computations will follow.

THEOREM 6.9. **[Structure Theorem]** *For any finite type $\mathbb{N}$-indexed persistence module $(V_\bullet, a_\bullet)$ over $\mathbb{F}$, there exists a set $\mathbf{Bar}(V_\bullet, a_\bullet)$ of integer pairs $0 \leq i \leq j \leq \infty$ (with $i \neq \infty$) and a function $\mu : \mathbf{Bar}(V_\bullet, a_\bullet) \to \mathbb{N}_{>0}$ to the nonzero natural numbers with the following property: there is a direct sum decomposition*

$$(V_\bullet, a_\bullet) \simeq \bigoplus_{[i,j]} (I_\bullet^{i,j}, c_\bullet^{i,j})^{\mu(i,j)}.$$

Here the indices $[i, j]$ range over elements of $\mathbf{Bar}(V_\bullet, a_\bullet)$. Moreover, this direct sum decomposition is unique (up to isomorphism of persistence modules).

PROOF. Since $(V_\bullet, a_\bullet)$ is of finite type, there is some $n \geq 0$ so that every $a_i : V_i \to V_{i+1}$ is an isomorphism for $i > n$. Consider the vector space $V = \bigoplus_0^n V_i$ and the linear map $t : V \to V$ sending each vector $v = (v_0, v_1, \ldots, v_n)$ to the shifted vector

$$t(v) = \big(0, a_0(v_0), a_1(v_1), \ldots, a_{n-1}(v_{n-1})\big).$$

This gives V the structure of a finitely generated $\mathbb{F}[t]$-module where $\mathbb{F}[t]$ is the polynomial ring over $\mathbb{F}$ in a single variable t. Since $\mathbb{F}[t]$ is a *principal ideal domain* whenever $\mathbb{F}$ is a field, every finitely generated $\mathbb{F}[t]$-module decomposes uniquely as a direct sum into two parts

$$V = F \oplus T,$$

where F is called *free* while T is *torsion*. Moreover, F is a direct sum of $\mathbb{F}[t]$-modules of the form $t^i \cdot \mathbb{F}[t]$ for some $i \geq 0$; each such free summand is isomorphic to an interval module of the form $(I_\bullet^{i,\infty}, c_\bullet^{i,\infty})$. Similarly, the torsion part T is a direct sum of modules of the form $t^i \cdot \mathbb{F}[t]/(t^j)$, i.e., a free module quotient by an ideal $(t^j) \lhd \mathbb{F}[t]$ with $0 \leq i < j$; each such summand is isomorphic to the interval module $(I_\bullet^{i,j}, c_\bullet^{i,j})$. These (free and torsion) interval modules might occur in the decomposition with any multiplicities ≥ 1, which are catalogued by the function μ. $\square$

While quite miraculous in its outcomes, this argument has two serious drawbacks arising from the fact that it invokes the classification of finitely generated $\mathbb{F}[t]$-modules. First, this proof strategy will not survive if we attempt something similar with $\mathbb{Z}[t]$-modules or even $\mathbb{F}[t_1, t_2]$-modules. Second, the *deus ex machina* nature of appealing to this classification renders life somewhat difficult for those who seek to understand the decomposition of $(V_\bullet, a_\bullet)$ on a more concrete level. There is no remedy for the first problem, but we can offer some solace to those afflicted by the second malady. The next Section contains a very concrete algorithm for computing interval-decompositions in the case of maximal interest to us, i.e., where $(V_\bullet, a_\bullet)$ arises from the homology groups of a filtered simplicial complex.

DEFINITION 6.10. For each ($\mathbb{N}$-indexed, finite type) persistence module $(V_\bullet, a_\bullet)$ over $\mathbb{F}$, the collection $\mathbf{Bar}(V_\bullet, a_\bullet)$ of intervals $[i, j]$ and their multiplicities $\mu(i, j) \geq 1$ (whose existence and uniqueness is guaranteed by Theorem 6.9) is called the **barcode** of $(V_\bullet, a_\bullet)$.

The content of Theorem 6.9 is that every finite type persistence module is uniquely determined up to isomorphism by the combinatorial data consisting of intervals $[i, j]$ in $\mathbf{Bar}(V_\bullet, a_\bullet)$ and their multiplicities $\mu(i, j)$. For brevity, we will denote multiplicities as superscripts, so $[1, 4]^3$ means that the bar $[1, 3]$ occurs with multiplicity $\mu(1, 4) = 3$ in a given barcode.

6.3 Algorithm (for Filtrations)

Let $\boldsymbol{F}_\bullet K$ be a filtered simplicial complex

$$\boldsymbol{F}_0 K \subset \boldsymbol{F}_1 K \subset \cdots \subset \boldsymbol{F}_{n-1} K \subset \boldsymbol{F}_n K = K,$$

and for each simplex σ of K let $b(\sigma)$ denote the smallest index i in $\{0, \ldots, n\}$ for which σ lies in $\boldsymbol{F}_i K$. Since each $\boldsymbol{F}_i K$ forms a subcomplex of K, it follows that b is an order preserving map on the simplices of K, i.e., $\sigma \leq \tau$ in K implies $b(\sigma) \leq b(\tau)$. In more prosaic terms, a simplex can only enter the filtration at index i if all of its faces are already present. Writing $g_{i \to j}$ for the inclusion map $\boldsymbol{F}_i K \hookrightarrow \boldsymbol{F}_j K$ for $i \leq j$, here we will describe an efficient algorithm which computes *all* the persistent homology groups $\mathbf{PH}_k g_{i \to j}(\boldsymbol{F}_\bullet K)$ at once by exploiting Theorem 6.9.

0. Setup: Order the simplices of K as $\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ so that σ precedes τ in this ordering when either on of the following conditions holds:

- we have $b(\sigma) \leq b(\tau)$, or
- we have $b(\sigma) = b(\tau)$ and σ is a face of τ in K.

Aside from these two constraints, the simplices of K may be ordered arbitrarily.

1. Input: The input to the algorithm is an $N \times N$ matrix $\boldsymbol{D}$ described as follows. For each pair (p, q) in $\{1, \ldots, N\}^2$, the entry of $\boldsymbol{D}$ in the p-th row and q-th column is given by

$$\boldsymbol{D}_{pq} = \begin{cases} \pm 1 & \text{if } \sigma_p \leq \sigma_q \text{ with } \dim \sigma_q - \dim \sigma_p = 1 \\ 0 & \text{otherwise} \end{cases}.$$

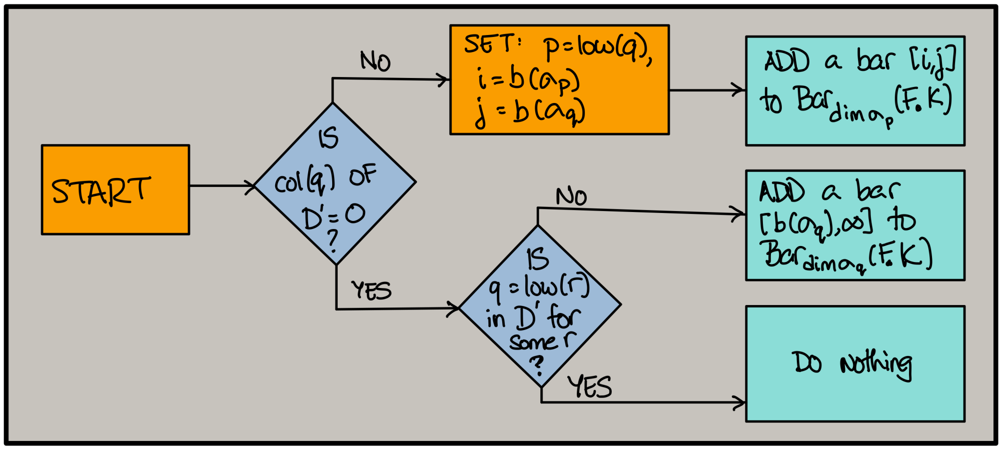
Here the sign ± 1 depends on an ordering of K's vertices; in particular, this is the same sign as the one used in the algebraic boundary operator of Definition 3.4. We will indicate the q-th column of $\boldsymbol{D}$ by $\text{col}(q)$ and write $\text{low}(q)$ to indicate the largest p satisfying $\boldsymbol{D}_{pq} \neq 0$, with the explicit understanding that $\text{low}(q) = 0$ whenever the $\text{col}(q)$ is entirely zero.

2. Procedure: The entire routine can be described with only six lines of pseudocode.

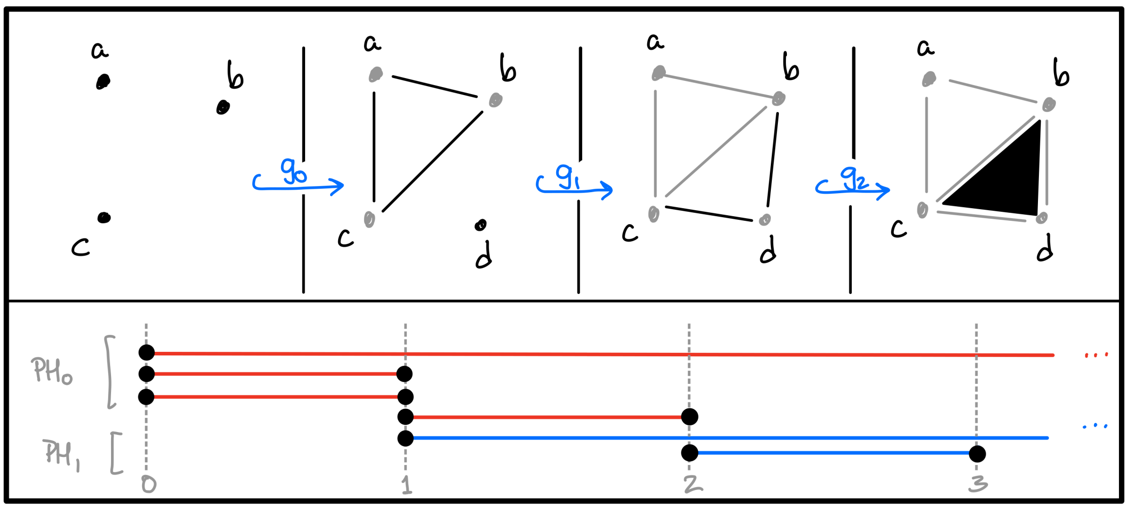
```
01 | For q = 1 to N
02 |     Set p = low(q)
03 |     While  some r < q satisfies low(r) = p ≠ 0
04 |         Add  (−D_pq/D_pr)·col(r) to col(q)
05 |     End While
06 | End For
```

3. Output: This procedure modifies the matrix $\boldsymbol{D}$ to produce a new matrix $\boldsymbol{D}'$ — this matrix $\boldsymbol{D}'$ is related to $\boldsymbol{D}$ by a change of basis since we only used column operations. In particular, lines 03-05 attempt to incrementally zero out the q-th column of $\boldsymbol{D}$ by adding preceding columns whose lowest nonzero entry coincides with that of $\text{col}(q)$. Thus, when the algorithm terminates, the p-th row of $\boldsymbol{D}'$ can be the lowest nonzero entry $\text{low}(q)$ of at most one column q — if there is such a q, then the entry $\boldsymbol{D}'_{pq}$ is said to form a *pivot* in the output matrix $\boldsymbol{D}'$.

4. The Barcodes: For each $k \geq 0$, let's write $\mathbf{Bar}_k(\boldsymbol{F}_\bullet K)$ to indicate the barcode of the persistence module obtained by taking the k-th homology groups of $\boldsymbol{F}_\bullet K$. We can read off such barcodes (and hence determine these persistence modules thanks to Theorem 6.9) by traversing the columns of $\boldsymbol{D}'$ and applying this handy flow-chart:



5. Example: When this algorithm is run on the filtration depicted in Section 1 (reproduced below), it will output the barcode $\{[0,\infty],[0,1]^2,[1,2]\}$ for 0-dimensional persistent homology and the barcode $\{[1,\infty],[2,3]\}$ for 1-dimensional persistent homology, perfectly capturing the evolution of connected components and loops at various stages in the filtration:



The starting point of the algorithm for this filtration is the following matrix $\boldsymbol{D}$ as described in the **Input** step above — all unlabelled entries are zero:

$$
D = \begin{array}{c} \\ a \\ b \\ c \\ d \\ ab \\ ac \\ bc \\ bd \\ cd \\ bcd \end{array}
\begin{array}{cccccccccc}
\overbrace{a \quad b \quad c}^{b=0} & & & \overbrace{d \quad ab \quad ac \quad bc}^{b=1} & & & & \overbrace{bd \quad cd}^{b=2} & & \overbrace{bcd}^{b=3}
\end{array}
$$

$$
\begin{array}{c|cccccccccc}
 & a & b & c & d & ab & ac & bc & bd & cd & bcd \\
\hline
a & & & & & -1 & -1 & & & & \\
b & & & & & 1 & & -1 & -1 & & \\
c & & & & & & 1 & 1 & & -1 & \\
d & & & & & & & & 1 & 1 & \\
ab & & & & & & & & & & \\
ac & & & & & & & & & & \\
bc & & & & & & & & & & 1 \\
bd & & & & & & & & & & -1 \\
cd & & & & & & & & & & 1 \\
bcd & & & & & & & & & &
\end{array}
$$

No operations are performed on ab's column, so the 1 in that column (in b's row) serves as a pivot. This pivot will contribute one of the two $[0, 1]$ bars in the 0-dimensional barcode of this filtration. The first interesting column operation occurs when the 1 in ac's column is used to clear out the 1 in bc's column (both corresponding to c's row). This changes the lowest entry in bc's column to the -1 in b's row, and we then use our pivot 1 in ab's column to cancel this new lowest entry. This will completely clear out bc's column, and contribute the $[1, \infty]$ bar in to the 1-dimensional barcode.

REMARK 6.11. Even on the small example described above, it is difficult to carry out the entire algorithm by hand. Fortunately, there are several good software packages available for computing persistent homology of filtered simplicial complexes arising in practice. In particular, one can find many implementations of this algorithm which will compute barcodes of Vietoris-Rips filtrations built around finite metric spaces (see Definition 1.15).

6.4 INTERLEAVING DISTANCE

Having witnessed the algebraic miracle of Theorem 6.9, we now turn to the geometric miracle, which takes the form of a **stability** result. Roughly, the set of finite type persistence modules admits the structure of a metric space, as does the set of barcodes; and with respect to the two chosen metrics, the assignment of a barcode to a module is an isometry. Here we will describe the desired metric on persistence modules after suitably upgrading them (and their barcodes) to be indexed by real numbers rather than natural numbers.

DEFINITION 6.12. An **$\mathbb{R}_+$-indexed persistence module** over $\mathbb{F}$ is a pair $(V_\bullet, a_\bullet)$ consisting of an $\mathbb{F}$-vector space V_t for each real number $t \geq 0$ and a linear map $a_{s\leq t} : V_s \to V_t$ for each pair $s \leq t$ of non-negative real numbers; these maps must satisfy

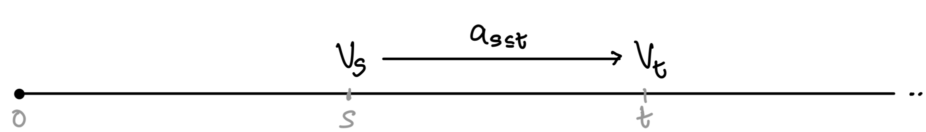
(1) $a_{t\leq t}$ is the identity map on V_t for each $t \geq 0$, and
(2) $a_{s\leq t} \circ a_{r\leq s} = a_{r\leq t}$ for every triple $0 \leq r \leq s \leq t$ of real numbers.

Put more succintly, these new persistence modules are functors of the form $(\mathbb{R}_+, \leq) \to \mathbf{Vect}_\mathbb{F}$ (see Definition 4.2). Here $(\mathbb{R}_+, \leq)$ is the category whose objects are all non-negative real numbers, with a unique morphism $s \to t$ whenever $s \leq t$; and the codomain is the usual category of (vector spaces, linear maps) over $\mathbb{F}$.

These persistence modules are more general than the $\mathbb{N}$-indexed ones from Definition 6.1: we can always replace an $\mathbb{N}$-indexed $(V_\bullet, a_\bullet)$ by an equivalent $\mathbb{R}_+$-indexed $(V'_\bullet, a'_\bullet)$ by interpolation as follows. Writing $\lfloor t \rfloor$ for the largest integer smaller than each t in $\mathbb{R}_+$, define

$$V'_t = V_{\lfloor t \rfloor} \quad \text{and} \quad a'_{s\leq t} = a_{\lfloor s \rfloor \to \lfloor t \rfloor}. \tag{4}$$

Henceforth, by persistence module we will mean the $\mathbb{R}_+$-indexed version defined above. For numerous reasons, it will be extremely convenient to visualize these as a continuum of of vector spaces living along a semi-infinite line segment connected by linear maps going from left to right, like so:



In order to guarantee barcodes for these new persistence modules a la Theorem 6.9, one must impose some finiteness constraints.

DEFINITION 6.13. A persistence module $(V_\bullet, a_\bullet)$ is called **tame** if two properties hold:

(1) the vector spaces V_t are finite-dimensional for all $t \geq 0$, and
(2) there are only finitely many $t \geq 0$, called *critical values*, for which the map $a_{t-\epsilon \leq t+\epsilon} : V_{t-\epsilon} \to V_{t+\epsilon}$ fails to be an isomorphism for arbitrarily small $\epsilon > 0$.

Tameness allows us to use Theorem 6.9 with impunity even with the $\mathbb{R}_+$-indexing — each tame persistence module $(V_\bullet, a_\bullet)$ can be reduced to a finite type $\mathbb{N}$-indexed persistence module $(V'_\bullet, a'_\bullet)$ as follows: let $0 \leq t_1 < t_2 < \cdots < t_n \leq \infty$ be the critical values of $(V_\bullet, a_\bullet)$ and set

$$V'_i = V_{t_i} \quad \text{and} \quad a'_i = a_{t_i \leq t_{i+1}}. \tag{5}$$

The barcode of $(V'_\bullet, a'_\bullet)$ can now be reinterpreted as the barcode of $(V_\bullet, a_\bullet)$ by sending each interval $[i,j]$ to the corresponding $[t_i, t_j]$. The *interval module* $(I_\bullet^{t_i,t_j}, c_\bullet^{t_i,t_j})$ supported on $[t_i, t_j]$ has the obvious definition:

$$I_t^{t_i,t_j} = \begin{cases} \mathbb{F} & t_i \leq t \leq t_j \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad c_{s\leq t}^{t_i,t_j} = \begin{cases} \mathrm{id}_{\mathbb{F}} & [s,t] \subset [t_i, t_j] \\ 0 & \text{otherwise} \end{cases}.$$

We have arrived at the following Corollary of Theorem 6.9; to fully appreciate its content, one should define (iso)morphisms and direct sums of tame persistence modules (as we did for their $\mathbb{N}$-indexed cousins).

COROLLARY 6.14. *For every tame persistence module $(V_\bullet, a_\bullet)$, there is a finite set $\mathbf{Bar}(V_\bullet, a_\bullet)$ of intervals of the form $[s,t] \subset \mathbb{R}_+$ (possibly with $t = \infty$) and a multiplicity $\mu : \mathbf{Bar}(V_\bullet, a_\bullet) \to \mathbb{N}_{>0}$ so that we have a unique direct sum decomposition into interval modules*

$$(V_\bullet, a_\bullet) \simeq \bigoplus_{[s,t]} (I_\bullet^{s,t}, c_\bullet^{s,t})^{\mu(s,t)},$$

with $[s,t]$ ranging over the intervals in $\mathbf{Bar}(V_\bullet, a_\bullet)$.

We now seek to measure distances between persistence modules. The following notion plays a central role.

DEFINITION 6.15. For each $\epsilon \geq 0$, an ϵ-**interleaving** between persistence modules $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ consists of two families of linear maps

$$\{\Phi_t : V_t \to W_{t+\epsilon} \mid t \geq 0\} \text{ and } \{\Psi_t : W_t \to V_{t+\epsilon} \mid t \geq 0\},$$

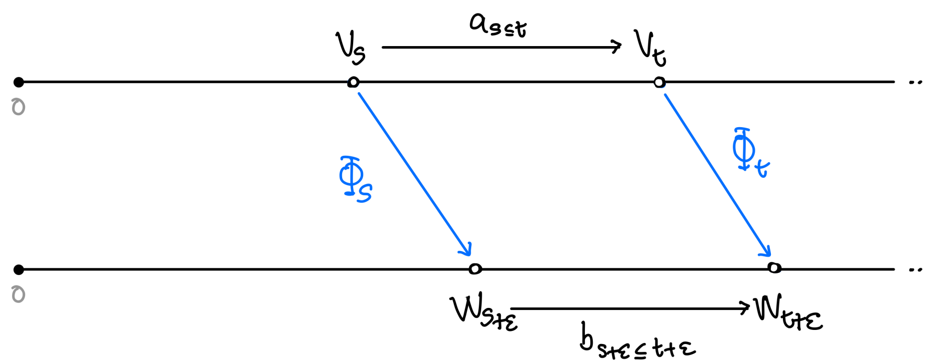
which satisfy four criteria. First, there are two *parallelogram relations*:

(1) for all $s \leq t$, we have $\Phi_t \circ a_{s\leq t} = b_{s+\epsilon \leq t+\epsilon} \circ \Phi_s$, and
(2) for all $s \leq t$, we have $\Psi_t \circ b_{s\leq t} = a_{s+\epsilon \leq t+\epsilon} \circ \Psi_s$.

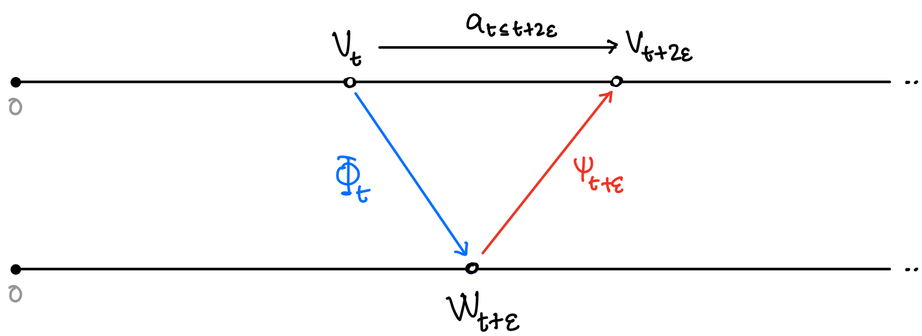
And second, there are two *triangle relations*:

(1) for all t, we have $\Psi_{t+\epsilon} \circ \Phi_t = a_{t\leq t+2\epsilon}$, and
(2) for all t, we have $\Phi_{t+\epsilon} \circ \Psi_t = b_{t\leq t+2\epsilon}$.

These four criteria might appear opaque at a first reading; the best method of acquiring an intuitive grasp on interleavings is to draw the commutative diagrams implied by the parallelogram and triangle relations. This will require us to visualize both $V_\bullet$ and $W_\bullet$ along line segments as suggested before, so that the maps Φ_t and Ψ_t connect each point $t \geq 0$ on one of these lines to the point $t + \epsilon$ on the other. Here, for instance, is the commuting diagram which represents the first parallelogram relation:



Of course, we have one such commuting diagram for *every* choice of $s \le t$. Similarly, here is an illustration of the first triangle relation (there is one such commuting triangle for every t).



It might also be helpful to verify that 0-interleavings are isomorphisms of persistence modules — this is one of the Exercises. Finally, here is the promised metric on persistence modules.

DEFINITION 6.16. The **interleaving distance** $\boldsymbol{d}_{\text{Int}}((V_\bullet, a_\bullet), (W_\bullet, b_\bullet))$ between persistence modules $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ is the infimum over all $\epsilon \ge 0$ for which there exists an ϵ-interleaving between them. If no such interleaving exists, then $\boldsymbol{d}_{\text{Int}}(V_\bullet, W_\bullet) = \infty$.

6.5 The Stability Theorem

The barcodes $\mathbf{Bar}(V_\bullet, a_\bullet)$ whose existence and uniqueness is guaranteed by Corollary 6.14 for each tame persistence module $(V_\bullet, a_\bullet)$ are finite multi-sets of intervals $[s,t] \subset \mathbb{R}_+ \cup \infty$. Here by multi-set we simply mean that each interval $[s,t]$ might have several copies within the barcode, the precise number being given by the function $\mu(s,t)$. Our next goal is to impose a metric on the set of all such multi-sets of intervals.

DEFINITION 6.17. For $\epsilon \ge 0$, an **ϵ-matching** between two multi-sets B and B' of intervals is a bijection $\rho : B_0 \to B_0'$ between a pair of multi-subsets $B_0 \subset B$ and $B_0' \subset B'$ subject to the following constraints:

(1) Every $[s,t]$ in $(B - B_0) \cup (B' - B_0')$ has length $t - s \le 2\epsilon$, and
(2) If $\rho[s,t] = [s',t']$ for some $[s,t]$ in B_0, then $|s - s'| \le \epsilon \ge |t - t'|$.

Thus, if ρ is an ϵ-matching between multi-sets B and B', then it must pair all intervals of length exceeding 2ϵ of B with those of B'. And moreover, if ρ pairs $[s,t]$ with $[s',t']$, then we can obtain s' and t' by perturbing s and t respectively by no more than ϵ:

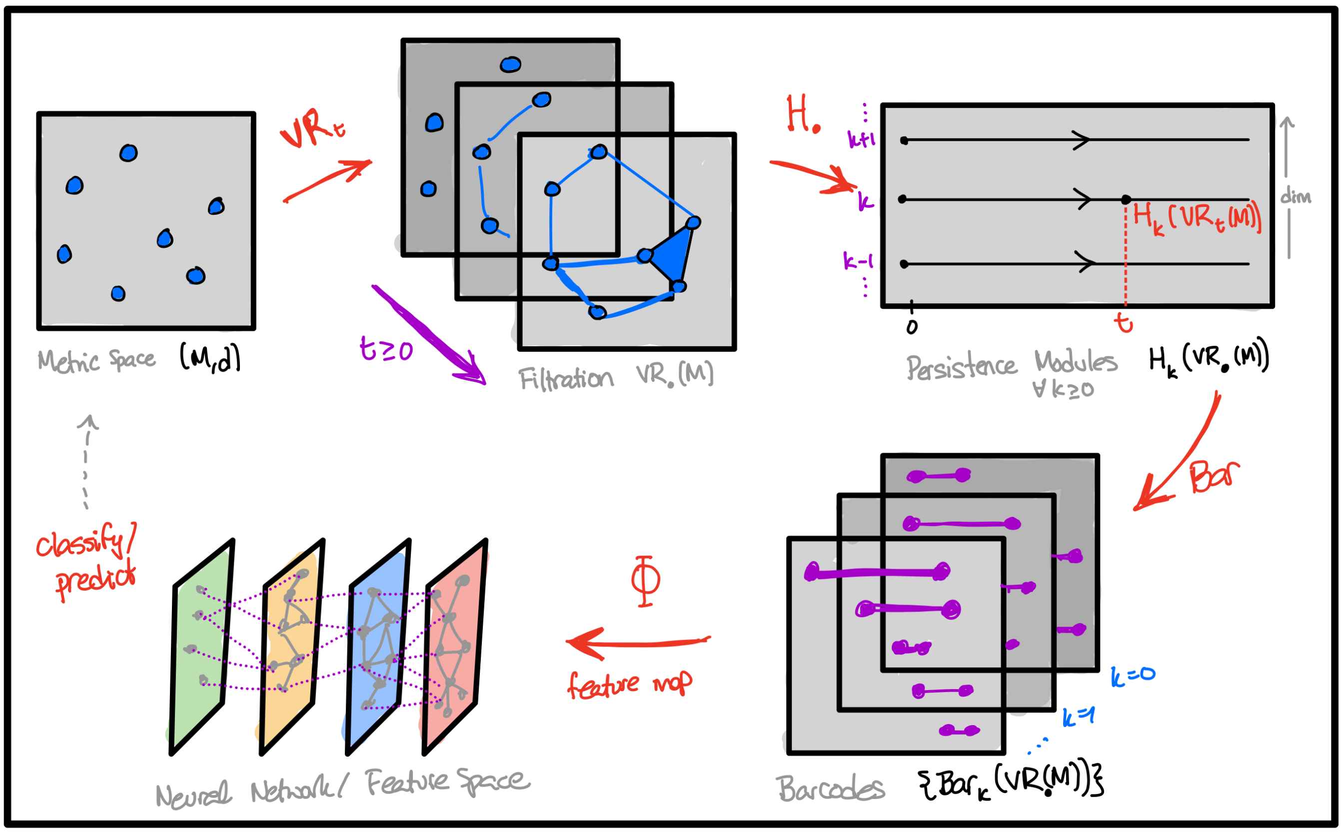
DEFINITION 6.18. The **bottleneck distance** $d_{\text{Bot}}(B, B')$ between multi-sets of intervals B and B' is the infimum over all $\epsilon \geq 0$ for which there exists an ϵ-matching between them.

Here is the geometric miracle of persistence modules.

THEOREM 6.19. **[Stability Theorem]** *For every pair $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ of tame persistence modules, we have*

$$d_{\text{Int}}((V_\bullet, a_\bullet), (W_\bullet, b_\bullet)) = d_{\text{Bot}}(\mathbf{Bar}(V_\bullet, a_\bullet), \mathbf{Bar}(W_\bullet, b_\bullet)).$$

Thus, the assignment of a barcode to a tame persistence module constitutes an isometry from the metric space of tame persistence modules (with interleaving distance) to the metric space of multi-sets of intervals (with bottleneck distance). All known proofs of the stability theorem are too technical to be included here[1]. The key advantage of the stability theorem is that it confers a certain geometric robustness to the following *topological data analysis* pipeline:



The first step describes the passage from a finite metric space to a filtered simplicial complex (as in Section 6 of Chapter 1). From there we compute persistent homology barcodes as described in Section 3 above. Since barcodes are combinatorial (rather than algebraic) objects, they can

[1]See Bauer and Lesnick's 2015 paper *Induced Matchings and the Algebraic Stability of Persistence Barcodes* for the most elementary proof known at present.

easily be vectorized and fed as input into neural networks or other statistical inference tools. The stability theorem enters the picture due to the following result.

PROPOSITION 6.20. *Let P and Q be two finite point-sets in $\mathbb{R}^n$ which are close in the following sense: there is some $\epsilon > 0$ so that*

(1) *there is a point of Q within distance ϵ of any point of P, and*
(2) *there is a point of P within distance ϵ of every point of Q.*

Then for each dimension $k \geq 0$, the k-th persistent homology modules of the Vietoris-Rips filtrations $\mathbf{VR}_\bullet(P)$ and $\mathbf{VR}_\bullet(Q)$ are 2ϵ-interleaved.

PROOF. Let $\alpha : P \to Q$ and $\beta : Q \to P$ be any pair of functions guaranteed by the ϵ-closeness of P and Q; thus, the Euclidean distance $\|p - \alpha(p)\|$ is no larger than ϵ for all p in P (and similarly for β). Now α induces simplicial maps $\{\alpha_t : \mathbf{VR}_t(P) \to \mathbf{VR}_{t+2\epsilon}(Q) \mid t \geq 0\}$ — to see why, note that if $\|p - p'\| \leq t$ then $\|\alpha(p) - \alpha(p')\| \leq t + 2\epsilon$ by the triangle inequality. Similarly, we get simplicial maps $\beta_t : \mathbf{VR}_t(Q) \to \mathbf{VR}_{t+2\epsilon}(P)$ for every $t \geq 0$. For each dimension $k \geq 0$, there are induced maps on homology $\mathbf{H}_k\alpha_t$ and $\mathbf{H}_k\beta_t$. We will now confirm that these induced maps $\mathbf{H}_k\alpha_t$ and $\mathbf{H}_k\beta_t$ satisfy the requirements of a 2ϵ-interleaving (Definition 6.15) between the persistence modules $\mathbf{PH}_k(\mathbf{VR}_\bullet P)$ and $\mathbf{PH}_k(\mathbf{VR}_\bullet Q)$.

1. Parallelogram Relations: For each $s \leq t$, let's denote the Vietoris-Rips inclusion maps as

$$i_{s\leq t} : \mathbf{VR}_s(P) \hookrightarrow \mathbf{VR}_t(P) \quad \text{and} \quad j_{s\leq t} : \mathbf{VR}_s(Q) \hookrightarrow \mathbf{VR}_t(Q).$$

By definition, we have $\alpha_t \circ i_{s\leq t} = j_{s+2\epsilon\leq t+2\epsilon} \circ \alpha_s$; now functoriality (i.e., Theorem 4.8) guarantees that the maps induced on k-th homology by α_s and α_t satisfy the parallelogram relation (see Definition 6.15).

$$\mathbf{H}_k\alpha_t \circ \mathbf{H}_k i_{s\leq t} = \mathbf{H}_k j_{s+2\epsilon\leq t+2\epsilon} \circ \mathbf{H}_k\alpha_s$$

An eerily similar argument establishes the parallelogram relation for $\mathbf{H}_k\beta_t$.

2. Triangle Relations: For each $t \geq 0$, note that the composite simplicial map

$$\beta_{t+2\epsilon} \circ \alpha_t : \mathbf{VR}_t(P) \to \mathbf{VR}_{t+4\epsilon}(P)$$

sends each vertex p to the vertex $p' = \beta \circ \alpha(p)$; by the triangle inequality we have $\|p - p'\| \leq 4\epsilon$. If $\sigma = (p_0, \ldots, p_m)$ is any m-simplex in $\mathbf{VR}_t(P)$, then the inclusion map $i_{t\leq t+4\epsilon}$ sends σ to σ, while the composite $\beta_{t+2\epsilon} \circ \alpha_t$ sends it to $\sigma' = (p'_0, \ldots, p'_m)$, with $p'_i = \beta \circ \alpha(p_i)$ for all i. It is easily confirmed that $\sigma \cup \sigma'$ is a simplex in $\mathbf{VR}_{t+4\epsilon}(P)$ by the triangle inequality. Thus, the simplicial maps $i_{t\leq t+4\epsilon}$ and $\beta_{t+2\epsilon} \circ \alpha_t$ are contiguous (in the sense of Corollary 2.9) and hence homotopic. By the homotopy invariance of homology (Theorem 4.24), their induced maps on homology coincide, and we obtain the desired triangle relation

$$\mathbf{H}_k\beta_{t+2\epsilon} \circ \mathbf{H}_k\alpha_t = \mathbf{H}_k i_{t\leq t+4\epsilon}.$$

A similar argument (with the roles of α and β interchanged) establishes the second triangle relation as well, and yields the desired result. □

As a consequence of the stability theorem, we see that for any $P, Q \subset \mathbb{R}^n$ satisfying the hypotheses of the above result, the k-th Vietoris-Rips persistent homology barcodes of P and Q must have the same number of sufficiently long bars, i.e., there is a bijection between bars of length $\geq 4\epsilon$ between the two barcodes in every homological dimension k. In this sense, the longer bars are stable to the sorts of perturbations which would replace P with Q. On the other hand, persistent homology is not stable to egregious outliers. In other words, if one obtains Q from P by adding just one point very far away from the existing points of P, then there is no relationship in general between the barcodes of P and those of Q.

EXERCISES

EXERCISE 6.1. Let $(V_\bullet, a_\bullet)$ and $(W_\bullet, b_\bullet)$ be $\mathbb{N}$-indexed persistence modules over a field $\mathbb{F}$. Show that for all $i \leq j$, there is an isomorphism

$$\mathbf{PH}_{i\to j}\big((V_\bullet, a_\bullet) \oplus (W_\bullet, b_\bullet)\big) \simeq \mathbf{PH}_{i\to j}(V_\bullet, a_\bullet) \oplus \mathbf{PH}_{i\to j}(W_\bullet, b_\bullet)$$

of persistent homology groups.

EXERCISE 6.2. Let $L \subset K$ be a two-step filtration of a simplicial complex K. Describe how to extract the dimension of the relative homology group $\mathbf{H}_k(K, L)$ for each $k \geq 0$ given the barcodes (with multiplicity) of this filtration.

EXERCISE 6.3. Let $\boldsymbol{F}_\bullet K$ be a filtration of a simplicial complex K. For each dimension $k \geq 0$ and filtration index i, describe how to compute the k-th Betti number of $\boldsymbol{F}_i K$ from the barcode $\mathbf{PH}_k(\boldsymbol{F}_\bullet K)$.

EXERCISE 6.4. Show that the interpolation of (4) produces an $\mathbb{R}_+$-indexed persistence module from an $\mathbb{N}$-indexed one.

EXERCISE 6.5. Describe a notion of morphisms which turn $\mathbb{R}_+$-indexed persistence modules into a category (if this is done correctly, the $\mathbb{N}$-indexed persistence modules will form a subcategory via (4)). What are the isomorphisms?

EXERCISE 6.6. Show that every $\mathbb{R}_+$-indexed interval module is tame.

EXERCISE 6.7. Show that sending a finite type $\mathbb{N}$-indexed persistence module $(V_\bullet, a_\bullet)$ to a tame $\mathbb{R}_+$-indexed one via (5), and then going back via (4), gives us $(V_\bullet, a_\bullet)$ back.

EXERCISE 6.8. Show that two ($\mathbb{R}_+$-indexed) persistence modules are isomorphic if and only if they admit a 0-interleaving.

EXERCISE 6.9. Draw commuting diagrams which represent the second parallelogram relation and the second triangle relation from Definition 6.15.

EXERCISE 6.10. Show that the interleaving distance satisfies the triangle inequality. [Hint: show that an ϵ-interleaving between $(U_\bullet, a_\bullet)$ and $(V_\bullet, b_\bullet)$ can always be combined with an ϵ'-interleaving between $(V_\bullet, b_\bullet)$ and $(W_\bullet, c_\bullet)$ to produce an $(\epsilon+\epsilon')$-interleaving between $(U_\bullet, a_\bullet)$ and $(W_\bullet, c_\bullet)$.]

EXERCISE 6.11. Let $a < a' < b < b'$ be four positive real numbers. What is the interleaving distance between the two $\mathbb{R}_+$-indexed interval modules $(I^{a,b}_\bullet, c^{a,b}_\bullet)$ and $(I^{a',b'}_\bullet, c^{a',b'}_\bullet)$?

EXERCISE 6.12. Show that the bottleneck distance satisfies the triangle inequality.

EXERCISE 6.13. State and prove a variant of Proposition 6.20 for Čech filtrations.

EXERCISE 6.14. Let $\mathscr{S}$ be a sheaf over a simplicial complex K and Σ an $\mathscr{S}$-compatible acyclic partial matching. Mimic the argument from Proposition 8.8 to show that the Morse complex of Σ with coefficients in $\mathscr{S}$ (see Definition 8.17) is a cochain complex.