

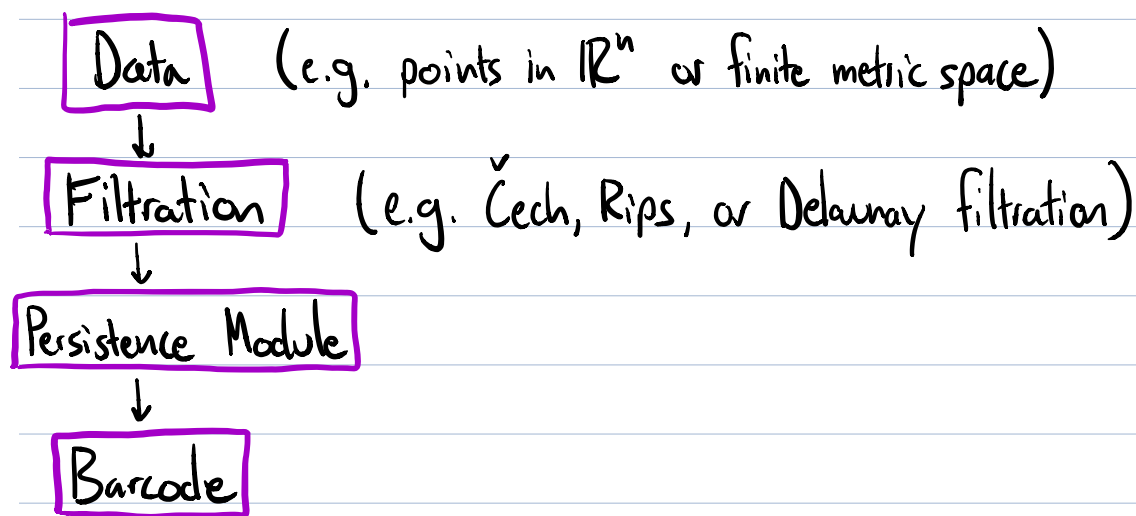
AMAT 584 Lecture 33, 4/17/20

Today: Persistent Homology

At various points in TDA I and II, I have mentioned the persistent homology pipeline, in varying levels of detail.

Here it is again:

Persistent Homology Pipeline



We've already discussed constructions for going from data to a filtration.

Now we discuss the rest of the pipeline.

Let's recall that a filtration (indexed by  $[0, \infty)$ ) is a collection of topological spaces

$$F = \{F_r\}_{r \in [0, \infty)} \text{ such that } F_r \subset F_s \text{ whenever } r \leq s.$$

This definition admits many variants:

- Simplicial complexes instead of topological spaces
- Filtrations indexed by  $\mathbb{N}$ ,  $\mathbb{Z}$ , or  $\mathbb{R}$  instead of  $[0, \infty)$ .

A filtration indexed by  $\mathbb{N}$  is just a sequence of spaces.

$$F_0 \subset F_1 \subset F_2 \subset \dots$$

### Persistence Modules

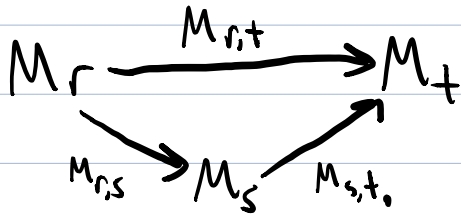
Let's fix a field  $F$ , say  $F = \mathbb{F}_2$ . A persistence module (indexed by  $[0, \infty)$ ) is a collection of vector spaces over  $F$

$$M = \{M_r\}_{r \in [0, \infty)} \text{ and linear maps}$$

$$\{M_{r,s}\}_{r \leq s} \text{ such that}$$

$$1) M_{r,r} = \text{Id}_{M_r} \quad \forall r \in [0, \infty)$$

$$2) M_{s,t} \circ M_{r,s} = M_{r,t} \quad \forall r \leq s \leq t, \text{ i.e. the following diagram commutes:}$$



As with filtrations, we can also talk about persistence modules indexed by  $\mathbb{N}$ ,  $\mathbb{Z}$ , or  $\mathbb{R}$ .

A persistence module indexed by  $\mathbb{N}$  is a sequence of vector spaces and linear maps

$$M_0 \xrightarrow{M_{0,1}} M_1 \xrightarrow{M_{1,2}} M_2 \xrightarrow{M_{2,3}} \dots$$

Note: I've only shown the maps between vector spaces at consecutive indices here. The remaining maps are given by composition, e.g.

$$M_{0,2} = M_{1,2} \circ M_{0,1}$$

$$\begin{array}{ccccccc}
 M_0 & \xrightarrow{M_{0,1}} & M_1 & \xrightarrow{M_{1,2}} & M_2 & \xrightarrow{M_{2,3}} & \dots \\
 & & & & \nearrow & & \\
 & & & & M_{0,2} & & 
 \end{array}$$

Given a filtration  $F$ , applying  $i^{\text{th}}$  homology to each space and each inclusion map gives a persistence module  $H_i(F)$ , as follows:

$$H_i(F)_r = H_i(F_r), \quad H_i(F)_{r,s} = H_i(F_r \hookrightarrow F_s)$$

inclusion map

$H_i(F)$  indeed satisfies conditions 1) and 2) in the definition of a persistence module; this follows immediately from the functoriality properties of homology.

The case of  $\mathbb{N}$ -indexed filtrations is simple:

$$F_0 \xrightarrow{j_0} F_1 \xrightarrow{j_1} F_2 \xrightarrow{j_2} \dots$$

For each  $i \geq 0$ , we get a persistence module

$$H_i(F_0) \xrightarrow{H_i(j_0)} H_i(F_1) \xrightarrow{H_i(j_1)} H_i(F_2) \xrightarrow{H_i(j_2)} \dots$$

It's a good idea to keep this  $\mathbb{N}$ -indexed case in mind. Although we work with  $[0, \infty)$ -indexed filtrations in TDA, these take only finitely many different values, and so are essentially  $\mathbb{N}$ -indexed filtrations, in a sense that can be made precise. (But I won't invest the time to do so.)

## Persistence Modules $\rightarrow$ Barcodes

A multiset is a set where elements can appear multiple times.  
e.g.,  $\{A, B, B, C\}$  is a multiset.

This can be defined formally, but I won't bother.

A barcode is a multiset of (non-empty) intervals in  $\mathbb{R}$ .  
e.g.,  $\{[1, 2), [0, 4), [0, 4), [3, 10)\}$  is a barcode.  
 $\{[0, 1), [0, 1], (0, 1], (0, 1)\}$  is also a barcode.

[In practice, each interval in a barcode arising in TDA is of the form  $[a, b)$ , but it is convenient for technical reasons to consider more general barcodes.]

In what follows, we work with  $[0, \infty)$ -indexing persistence modules, though everything we will say adapts to  $\mathbb{N}$ ,  $\mathbb{Z}$ , or  $\mathbb{R}$  indices.

Def: A compatible set of bases  $\mathcal{B}$  for a persistence module  $M$  is a choice of basis  $B_r$  for each vector space  $M_r$  of  $M$ , such that

- 1) if  $\forall r \leq s$  and  $b \in B_r$ , then either  $M_{r,s}(b) \in B_s$  or  $M_{r,s}(b) = 0$ .
- 2) If  $b_1 \neq b_2 \in B_r$  and  $M_{r,s}(b_1) \neq \vec{0}$ , then  $M_{r,s}(b_1) \neq M_{r,s}(b_2)$ .

Given a compatible set of bases  $\mathcal{B}$  for  $M$ , we can construct a barcode. The main idea is that the basis elements mapping to one another

"chain together" into intervals. Here are the details!

Consider the set  $\mathbb{L}B = \{(b, r) \mid b \in B_r\}$ . Let  $\rho: \mathbb{L}B \rightarrow [0, \infty)$  be given by  $\rho(b, r) = r$ .

Define an equivalence relation  $\sim$  on  $\mathbb{L}B$  by  $(b, r) \sim (b', r')$  iff  $M_{r', r}(b) = b'$  or  $M_{r, r'}(b') = b$ .

(It's straightforward to check that this is an equivalence relation.)

Proposition:  $\text{Barc}(B) = \{\rho(E) \mid E \text{ an equivalence class of } \sim\}$  is a barcode.

We say a persistence module  $M$  is pointwise finite dimensional (p.f.d. for short) if each vector space  $M_r$  is finite dimensional.

Theorem: If  $M$  is p.f.d. then there exists a compatible basis  $B$  for  $M$ . Moreover,  $\text{Barc}(B)$  is independent of the choice of  $B$ . Thus we obtain a well defined barcode  $\text{Barc}(M)$ .