ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Theoretical Computer Science
Patrick Schnider

Introduction to Topological Data Analysis **Scribe Notes 18** **FS23**

Scribe notes by Simon Weber. Please contact me for corrections.
Lecture date: May 4, 2023
Last update: Thursday 4th May, 2023, 17:18

Interval Persistence Modules

We consider persistence modules over $\mathbb{R}$ of vector spaces over some field $\mathbb{F}$.

In today's lecture we look at some special persistence modules, called interval modules.

Definition 1. *A interval module* $\mathbb{I}[b, d]$ *is an persistence module*

$$V_a = \begin{cases} \mathbb{F} & \text{if } a \in [b, d], \\ 0 & \text{otherwise.} \end{cases} \qquad \text{and} \qquad v_{a,a'} = \begin{cases} \mathrm{id} & b \le a \le a' \le d, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we can define interval modules on open and clopen intervals, denoted by $\mathbb{I}(b, d)$, $\mathbb{I}(b, d]$, and $\mathbb{I}[b, d)$. We write $\mathbb{I}\langle b, d\rangle$ to include all four of these types.

For an interval module we can easily talk about birth and death as we did in persistent homology. If we have a persistent homology module that is (isomorphic to) an interval module, the birth and death correspond to the boundaries b, d of the interval.

Definition 2. *A persistence module* $\mathbb{U}$ *is called* pointwise finite dimensional (p.f.d.) *if for all* $a \in \mathbb{R}$, U_a *has finite dimension.*

Note that all p.f.d. persistence modules are also q-tame.

Definition 3. *Given two persistence modules* $\mathbb{U}, \mathbb{V}$, *we define their direct sum* $\mathbb{U} \oplus \mathbb{V}$ *by* $(U \oplus V)_a = U_a \oplus V_a$ *and* $(u \oplus v)_{a,a'} = u_{a,a'} \oplus v_{a,a'}$.

Here, the direct sum of maps just means applying the respective maps component-wise.

Proposition 4. *If* $\mathbb{U}_1, \mathbb{U}_2$ *are* ϵ*-interleaved, and* $\mathbb{V}_1, \mathbb{V}_2$ *are* δ*-interleaved, then* $\mathbb{U}_1 \oplus \mathbb{V}_1$ *and* $\mathbb{U}_2 \oplus \mathbb{V}_2$ *are* $\max\{\epsilon, \delta\}$*-interleaved.*

Proof. W.l.o.g. $\epsilon \geq \delta$, so we need to show that they are ϵ-interleaved. Recall that if two persistence modules are δ-interleaved, they are also ϵ-interleaved. Let φ^u, ψ^u be (series of) functions showing that $\mathbb{U}_1, \mathbb{U}_2$ are ϵ-interleaved. Similarly, let φ^v, ψ^v be (series of) functions showing that $\mathbb{V}_1, \mathbb{V}_2$ are ϵ-interleaved. Then, $\varphi^u \oplus \varphi^v, \psi^u \oplus \psi^v$ show that $\mathbb{U}_1 \oplus \mathbb{V}_1$ and $\mathbb{U}_2 \oplus \mathbb{V}_2$ are ϵ-interleaved. □

If we now have a direct sum of interval modules, we can still nicely talk about birth and death: We just look at each interval module in isolation. The following theorem shows that surprisingly most persistence modules can be expressed as direct sums of interval modules.

Theorem 5 (Structure theorem). *Any p.f.d. persistence module decomposes uniquely into interval modules, i.e., we have*

$$\mathbb{U} \cong \bigoplus_{i \in I} \mathbb{I}\langle b_i, d_i \rangle.$$

The intervals $\langle b_i, d_i \rangle$ are exactly the barcodes if $\mathbb{U}$ is a persistent homology module.

Note that unless we have some additional tame-ness condition on $\mathbb{U}$, I is not guaranteed to be finite.

Recall that when we talked about persistent homology, we said that there is some consistent global choice of basis for persistent homology groups. That was a consequence of the structure theorem.

Proposition 6. *Consider two interval modules $\mathbb{I}_1 = \mathbb{I}\langle b_1, d_1 \rangle$ and $\mathbb{I}_2 = \mathbb{I}\langle b_2, d_2 \rangle$. Then, $d_I(\mathbb{I}_1, \mathbb{I}_2) = d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$.*

Proof. (This proof has not been shown in the lecture and is not relevant for the exam.) To prove that $d_I(\mathbb{I}_1, \mathbb{I}_2) \geq d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$, we show that every upper bound on d_I is also an upper bound on d_b: assume that we have maps φ, ψ showing that the two modules are ϵ-interleaved. Then, consider $\psi_{a+\epsilon} \circ \varphi_a = v^1_{a,a+2\epsilon}$, equality holding because φ, ψ certify ϵ-interleaving. Consider $a \in \langle b_1, d_1 \rangle$.

Case 1: $v^1_{a,a+2\epsilon} = 0$ for all $a \in \langle b_1, d_1 \rangle$. Then, $d_1 - b_1 < 2\epsilon$, and the (infinity-norm) distance of (b_1, d_1) to the diagonal is less than ϵ.

Case 2: $v^1_{a,a+2\epsilon} = \mathrm{id}$ for some $a \in \langle b_1, d_1 \rangle$. Then, $d_1 - b_1 \geq 2\epsilon$. Furthermore, we have $\varphi_a(\mathbb{F}) = \mathbb{F}$ for all $a \in \langle b_1, d_1 - 2\epsilon \rangle$. So, for these a, we must also have $a + \epsilon \in \langle b_2, d_2 \rangle$. This tells us that $\langle b_2, d_2 \rangle$ must "cover" a large part of $\langle b_1, d_1 \rangle$, namely we get $b_2 \leq b_1 + \epsilon$, and $d_2 \geq d_1 - \epsilon$. We can now see that $|b_2 - b_1| \leq \epsilon$ and $|d_2 - d_1| \leq \epsilon$: to violate this, $\langle b_2, d_2 \rangle$ would have to be a larger interval than $\langle b_1, d_1 \rangle$ (in particular, it would be longer than 2ϵ), and we could thus exchange their roles and get that $b_1 \leq b_2 + \epsilon$ and $d_1 \geq d_2 - \epsilon$. From this, we get that $d_\infty((b_1, d_1), (b_2, d_2)) \leq \epsilon$, and thus get the bound on d_b.

We now prove the other direction, $d_I(\mathbb{I}_1, \mathbb{I}_2) \leq d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$. To see this, we show that from every matching whose longest edge is ϵ, we get an ϵ-interleaving.

Case 1: The two off-diagonal points are matched to the diagonal. Then, we get that $d_i - b_i \leq 2\epsilon$ for both of them, and thus for all $\epsilon' > \epsilon$, $\mathbb{I}_1$ and $\mathbb{I}_2$ are ϵ'-interleaved with $\varphi, \psi = 0$. Thus, $d_I \leq \epsilon$.

Case 2: The points are matched with each other. Then, $|b_2 - b_1| \leq \epsilon$ and $|d_2 - d_1| \leq \epsilon$. Taking $\varphi, \psi = \mathrm{id}$ we can see that $\mathbb{I}_1$ and $\mathbb{I}_2$ are ϵ-interleaved. Thus, $d_I \leq \epsilon$. □

Corollary 7. *Let $\mathbb{U}, \mathbb{V}$ be p.f.d. persistence modules. Then,* $d_I(\mathbb{U}, \mathbb{V}) \leq d_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V})$.

Proof. We apply the structure theorem to write $\mathbb{U} = \bigoplus_{i \in I} \mathbb{I}\langle b_i, d_i \rangle \oplus \bigoplus_{j \in J} 0$ and $\mathbb{V} = \bigoplus_{j \in J} \mathbb{I}\langle b_j, d_j \rangle \oplus \bigoplus_{i \in I} 0$. From the Bottleneck matching we get a matching between parts making up $\mathbb{U}$ and $\mathbb{V}$. Since the Bottleneck distance is the maximum length of any edge, we have $d_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V}) \geq d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2) = d_I(\mathbb{I}_1, \mathbb{I}_2)$ for every two interval modules that were matched together, where we used Theorem 6. Finally, we use Theorem 4 to get the desired statement. □