The Bottleneck distance and the Interleaving distance

Sigurd Gaukstad

October 6, 2021

Contents

1. What we will learn in this lecture

(1) A recap (Jupyter Notebook)
(2) The Bottleneck distance of barcodes/percistent diagrams and stability
(3) Interleaving distance of Modules and stability
(4) The Isometry Theorem
(5) Summary

The main sources is [Bot] and [**Chapter 3**; Oud15].

2. The bottleneck distance

If we have two point clouds, we get two persistence diagrams. A natural question to ask is how we can compare two persistence diagrams, when are they equal, and if not, how far from equal are they. Furthermore does persistence diagrams give a good invariant for point clouds? If two point clouds are almost the same, we would like to have similar persistence diagrams, but at the same time, persistence diagrams should not be so coarse of an invariant that it does not separate point clouds not begin similar. We aim to make this precise by defining a metric *the Bottleneck distance* on the persistence diagrams.

NOTATION 0.1. We will let $\mathcal{C}$ and $\mathcal{D}$ denote *multisets* of intervals $\langle a, b\rangle \subset \mathbb{R}$, where the brackets $\langle a, b\rangle$ can mean the closed, open or half closed/half open interval between a and b. Multiset simply means a set which we allow for multiple instances for each of its elements. For example, $\mathcal{C} = \{[0,1], [0,1], (6,10), [5.4, -113)\}$. The thing to have in mind is barcodes or persistence diagrams.

DEFINITION 0.2. We define a *matching* between $\mathcal{C}$ and $\mathcal{D}$ to be a collection of pairs $\chi = \{(I, J) \in \mathcal{C} \times \mathcal{D}\}$, where the intervals I and J can occur in at most one pair. If $(I, J) \in \chi$, then we say that I is *matched* to J, otherwise if I does not belong to any pair in χ, we say that I is *unmatched*.

We remark that a matching between multisets $\mathcal{C}$ and $\mathcal{D}$ have a equivalent description as a bijection between a subset of $\mathcal{C}$ and a subset of $\mathcal{D}$. In this sense, there exist a notion of composition of matchings, by restricting the bijections sufficently.

EXAMPLE 0.3. Let $\mathcal{C} = \{(0,1), [0,1), (0,1)\}$ and $\mathcal{D} = \{(-\pi, 0], (0,2)\}$, then $\chi = \{((0,1), (-\pi, 0])\}$ is a matching between $\mathcal{C}$ and $\mathcal{D}$ such that $(0,1)$ is matched to $(-\pi, 0]$ and $[0,1), (0,1) \in \mathcal{C}$ and $(0,2) \in \mathcal{D}$ are unmatched.

DEFINITION 0.4. Let $\mathcal{C}$ and $\mathcal{D}$ be multisets of intervals in $\mathbb{R}$ with a matching χ. Then we define a function $c\colon \chi \to \mathbb{R}$, called the *cost*, which maps a pair $(I, J) = (\langle a, b\rangle, \langle c, d\rangle)$, to

$$(\langle a, b\rangle, \langle c, d\rangle) \mapsto \max(|c - a|, |d - b|)$$

We will define the cost of an interval $I = \langle a, b\rangle$ to be

$$c(I) := (b-a)/2$$

Furthermore we define the *cost* of matching χ as

$$c(\chi) := \max\Big(\sup_{(I,J)\in\chi} c(I,J), \sup_{\text{unmatched} I\in C\cup\mathcal{D}} c(I)\Big)$$

For $\epsilon \in \mathbb{R}$, we say that χ is an *ϵ-matching* if $c(\chi) \leq \epsilon$

We are now ready to define the *bottleneck distance*

Definition 0.5. The **bottleneck distance** between C and $\mathcal{D}$, two multisets, is

$$d_B(C,\mathcal{D}) := \inf\{c(\chi) \mid \chi \text{ i a matching between } C \text{ and } \mathcal{D}\}$$

Remark 0.6. This actually defines a metric on the set of multisets:

(1) $d_B(C,\mathcal{D}) = 0$ if and only if $C = \mathcal{D}$
(2) $d_B(C,\mathcal{D}) = d_B(\mathcal{D},C)$ (Symmetry)
Both of the above should be clear from the definition.
(3) $d_B(C,\mathcal{D}) \leq d_(C,\mathcal{E}) + d_B(\mathcal{E},\mathcal{D})$ (Triangle identity) This follows from:

Lemma 0.7. [**Lemma 9.3**; Bot] *If there exists an ϵ-matching between C and $\mathcal{D}$, and an $\tilde{\epsilon}$-matching between $\mathcal{D}$ and $\mathcal{E}$, then the composition of matching yield an $(\epsilon+\tilde{\epsilon})$-matching between C and $\mathcal{D}$.*

2.1. Q and A. How fast can we compute the Bottleneck distance?
As mentioned in Botnans notes, if we let C and $\mathcal{D}$ be finite multisets and $n := |C| + |\mathcal{D}|$. Then the state-of-the-art algorithms compute the Bottleneck distance $d_B(C,\mathcal{D})$ in $O(n^{1.5}\log n)$, which is acceptable.

Why is it called "Bottleneck"?
I don't know for sure, but maybe because, there are a lot of potential matchings χ, and you check one at a time, making a bottleneck as in a traffic jam.

What is good about the Bottleneck distance?

(1) It satisfy being a metric (When we restrict to finite multisets, as we do).
(2) It is *stable*!

2.2. Stability. With *stability* we mean the following: a small change in the input filtration, yields a small change of its persistence diagram in the Bottleneck distance. For any topological space X equipped with two real valued functions $f, g\colon X \to \mathbb{R}$. To these functions we have a family of $\mathbb{R}$-modules. which we assume to be finite dimensional, defined by

$$M_t^f := H_i(f^{-1}(-\infty,t]) \textbf{ and } H_t^g := H_i(g^{-1}(-\infty,t])$$

Recall that $M^f = \{M_t^f\}_{t\in\mathbb{R}}$ is a persistence module, with maps induced from inclusions $f^{-1}(-\infty,t] \hookrightarrow f^{-1}(-\infty,s]$ for $t \leq s$. Also recall that from a persistence module M^f, there is a corresponding barcode $B(M^f)$, from last week talk or

[**Theorem 6.18**; Bot]. Then one can show the following theorem, which we will call for the **Stability Theorem**:

THEOREM 0.8. *[**Theorem 9.6**; Bot]*

$$d_B(B(M^f), B(M^g)) \leq ||f - g||_\infty$$

Where $||f - g||_\infty := \inf\{C \geq 0 \colon |f(x) - g(x)| \leq C$ *for almost every* $x\}$ *is the L-infinity norm.*

There are other different stability theorems for the bottleneck metric, for different settings. We have one for the Vietoris–Rips filtrations, where $H_i(VR(P))$ denotes the $\mathbb{R}$-indexed persistence module:

THEOREM 0.9. *[**Theorem 9.7**; Bot] For any bijection of finite subset of* $\mathbb{R}^d$ $\sigma \colon P \to Q$ *such that* $||p - \sigma(p)|| \leq \epsilon$ *for all* $p \in P$*, then*

$$d_B(B(H_i(VR(P))), B(H_i(VR(Q)))) \leq \epsilon$$

And we have one for the Cech filtration: Let $d_P \colon \mathbb{R}^d \to \mathbb{R}$ be defined as

$$d_P(x) := \min_{y \in P} ||x - y||$$

i.e. the distance of the point $x \in \mathbb{R}$ to a closest point $y \in P$ in the point cloud. Then we can define $P_r := d_P^{-1}(-\infty, r]$. I.e. the points in $\mathbb{R}^d$ that closer then r to the point cloud P.

THEOREM 0.10. *[**Theorem 9.8**; Bot] Let* P *and* Q *be point clouds in* $\mathbb{R}^d$*. If there exist a* $\epsilon \geq 0$ *such that* $P \subset Q_\epsilon$ *and* $Q \subset P_\epsilon$*, i.e. the point clouds are in the sense as above closer then* ϵ *from each other, then*

$$d_B(B(H_i(Cech(P))), B(H_i(Cech(Q)))) \leq \epsilon$$

Upshots from the stability theorems:
If one compares two identical point clouds or barcodes, then changing one of them by a certain amount is reflected in the metric. We will therefore say that the Bottleneck distance is stable.

In plain English, this stability says that if two sets $X, Y \subset \mathbb{R}^n$ are Hausdorff close, then their barcodes are Bottleneck close.

What is bad about the Bottleneck distance?

(1) Although it is table against wiggeling the point cloude, it is not stable with reguards of adding points to the cloude. For instance if one adds a point in the middle of a circle of points.
(2) It seems that it is hard to generalize the Bottleneck distance beyond 1-persistence modules, i.e. to multi-persistence modules, as it relies on barcodes. What about QR-codes? This leads us to the discussion of another metric, called *Interleaving distance.*

As the Bottleneck distance only sensitive to the maximum over a matching, there are other metrics such as **degree q Wasserstein distance** that is more used in data analysis, although they do not exhibit the same type of stability as the bottleneck distance. However, the Wasserstein distance coincide with the Bottleneck distance for $p = \infty$.

3. Interleaving distance of persistence modules

We now want to make a metric on the persistence modules themself, without going to barcodes. For a moment we will assume that our peristence modules are indexed over the intergers $\mathbb{Z}$. We define a 0-**interleaving** of persictence modules M and N, as morphisms

$$\begin{tikzcd} M_t \arrow[r, bend left] & N_t \arrow[l, bend left] \end{tikzcd}$$

for every $t \in \mathbb{Z}$, such that the following diagram commute:

$$\begin{tikzcd} \dots \arrow[r] & M_{i-1} \arrow[r] \arrow[d, bend left] & M_i \arrow[r] \arrow[d, bend left] & M_{i+1} \arrow[r] \arrow[d, bend left] & \dots \\ \dots \arrow[r] & N_{i-1} \arrow[r] \arrow[u, bend left] & N_i \arrow[r] \arrow[u, bend left] & N_{i+1} \arrow[r] \arrow[u, bend left] & \dots \end{tikzcd}$$

In other words the zero inter leaving is simply a collection of isomorphims

$$\begin{tikzcd} M_t \arrow[r, bend left] & N_t \arrow[l, bend left] \end{tikzcd}$$

and therefore an isomorphism of persistence modules. Similarly, we define a 1-**interleaving** as a collection of morphisms

$$M_t \longrightarrow N_{t+1} \qquad N_t \longrightarrow M_{t+1}$$

for every $t \in \mathbb{Z}$ such that the following diagram commute:

$$\begin{tikzcd} \dots \arrow[r] \arrow[dr] & M_{i-1} \arrow[r] \arrow[dr] & M_i \arrow[r] \arrow[dr] & M_{i+1} \arrow[r] \arrow[dr] & \dots \\ \dots \arrow[r] \arrow[ur] & N_{i-1} \arrow[r] \arrow[ur] & N_i \arrow[r] \arrow[ur] & N_{i+1} \arrow[r] \arrow[ur] & \dots \end{tikzcd}$$

This is not an isomorphism of persistence modules. More generally, we can define a z-interleaving of $\mathbb{Z}$-indexed persistence modules in the following way.

DEFINITION 0.11. Let M, N be $\mathbb{Z}$-indexed persistence modules, then a z-interleaving is a collection of morphisms

$$M_t \longrightarrow N_{t+z} \qquad N_t \longrightarrow M_{t+z}$$

making the obvious diagram commute.

Of course, we don't want to restrict our self to $\mathbb{Z}$-indexed modules, as all our examples from today's lecture are $\mathbb{R}$-indexed. First we need the notion of *shift*-module.

Definition 0.12. Let M and N be $\mathbb{R}$-indexed persistence modules. Let $\epsilon \in \mathbb{R}$. We define the **ϵ-shit** of a persistence module M to be the module M^ϵ, defined by $M_t^\epsilon := M^{t+\epsilon}$, with the same maps as M. Moreover for any $f\colon M \to N$, we get an induced map $f^\epsilon\colon M^\epsilon \to N^\epsilon$.

Now we are able to define the more general ϵ-*interleaving.*

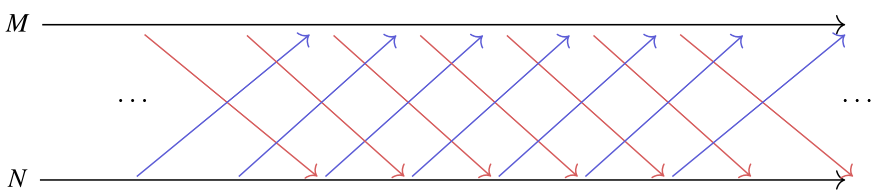
Definition 0.13. Let M and N be $\mathbb{R}$-indexed persistence modules. Given an $\epsilon \in [0, \infty)$, an **ϵ-interleaving** (Blande, Stokke, "Interfoliere, å sette inn blanke blad mellom de trykte bladene i en bok. ") between M and N, is a pair of morphisms of persistence modules

$$M \xrightarrow{\psi} N^\epsilon \qquad \text{and} \qquad N \xrightarrow{\phi} M^\epsilon$$

such that the following two diagrams commute for any $t \in \mathbb{R}$:

$$\begin{tikzcd}
M_t \arrow[rr, "M(t\leq T+2\ \epsilon)"] \arrow[dr, "\psi_t"'] & & M_{t+2\epsilon} \\
& N_{t+\epsilon} \arrow[ur, "\phi_{t+\epsilon}"'] &
\end{tikzcd}
\qquad \text{and} \qquad
\begin{tikzcd}
& M_{t+\epsilon} \arrow[dr, "\psi_{t+\epsilon}"] & \\
N_t \arrow[ur, "\phi_t"] \arrow[rr, "N(t\leq t+2\epsilon)"'] & & N_{t+2\epsilon}
\end{tikzcd}$$

We remark that this is exactly the same type of diagram as we got for $\mathbb{Z}$-indexed persistence modules. Considering M and as a copies of the real line, with each number representing the corresponding module we get the following diagram with a continuum of arrows between the modules, with the same slope, decided by ϵ, making the diagram commute:



We are now ready to define **interleaving metric** between $\mathbb{R}$-indeced persistence modules.

DEFINITION 0.14. Let M and N be $\mathbb{R}$-indexed persistence modules. The the **interleaving distance** between M and N is defined as

$$d_I(M, N) := \inf\{\epsilon\colon \text{there exist a } \epsilon\text{interleaving between } M \text{ and } N\}$$

We should think of this metric as measuring how far two persistence modules are from being isomorphic. Knowing that a 0-interleaving is an isomorphism.

REMARK 0.15. This actually defines a metric. The two first conditions are immediate. And the third is left as a fun exercise/see [**Lemma 9.11**; Bot].

REMARK 0.16. The interleaving distance is universal in the following sense, if d is any other metric of persistence modules, then $d \leq d_i$.

3.1. The isometry theorem. One can ask themselves if this is a good metric, and the answer is yes. The reason is the following theorem:

THEOREM 0.17. *(Isometry theorem) Let M and N be $\mathbb{R}$-indexed persistence modules. Then*

$$d_B(B(M), B(N)) = d_I(M, N)$$

PROOF. This will be discussed next week. □

How to interpret this theorem? Knowing that

$$d_B(B(M), B(N)) \leq d_I(M, N)$$

We have on the right side, a barcode-less definition. We can say that barcodes, or persitent diagrams, are stable signatures for persitence moduels. The converse

$$d_B(B(M), B(N)) \geq d_I(M, N)$$

tells us that barcodes, or persitent diagrams are informative.

REMARK 0.18. A multiparameter persisent module is a functor $M\colon R^d \to F-$Mod. Where $R \subset \mathbb{R}$ is a subset, and R^d has the poset structure. For instance R^2, then $(i, j) \leq (p, q)$ if and only if $i \leq p$ and $j \leq q$. How can such a module arrise in TDA? For instance if $f : Q \to \mathbb{R}$ is a function from a finite metric space Q. We know that looking at $f^{-1}(-\infty, \delta_i]$, gives a filtration of our space Q. Suppose that $\delta_1, \ldots \delta_n$, is a set of thresholds were interesting things happening, for instance a hole disappears or appears. Suppose also that $\epsilon_1, \ldots, \epsilon_m$ are some interesting choices for ϵ such that $VR\epsilon_i(f^{-1}(\infty, \delta_j)])$ is an interesting filtration of the Vietori Rips complex for every δ_j. Taking homology (with coefficients in F) gives a multiparameter persistent homology, or Quiver over F, with the induced maps from inclusions and

$$M_{i,j} = H_l(VR\epsilon_i(f^{-1}(\infty, \delta_j)]))$$

for every l. We note that this Quiver will look like a grid.
The point however is that interleaving distance of modules, still makes sense in this case, giving the interleaving distance an edge over the Bottleneck distance

REMARK 0.19. Computationally, interleaving is not so nice. Interleaving distance is NP-hard in general by [**Theorem 2**; BB18].

4. Summary

We set out to find a way to say if two persistence modules are the same, and if not, how far appear are they. Knowing that persistence modules is in bijection with barcodes/persistence diagrams, We defined a metric, the *Bottleck distance* on the latter. It turned out to satisfy begin a metric. Furthermore it satified a some *stability* results and is not too bad computationally. One of the problems with this, is that it does not generalize very well to multiparameter persistence modules. It relied on the Barcodes. We therefore defined a metric on the persitence modules themself, called the *interleaving metric*, it too satified being a metric. This is much easier to generalize to multiparameter persistence modules, however computationally, this is harder. We also saw that it is stable as, it agreed with the *Bottlneck* distance on the 1-dim persistence modules via the *Isometry Theorem*. The *Isometry theorem* showed us that *Barcodes* are *informative* and *stable*.

Bibliography

[BB18] H. B. Bjerkevik and M. B. Botnan. *Computational Complexity of the Interleaving Distance*. 2018.

[Bot] M. B. Botnan. *Topological Data Analysis Spring 2020*. https://www.few.vu.nl/~botnan/lecture_notes.pdf.

[Oud15] S. Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. Mathematical Surveys and Monographs 209. American Mathematical Society, 2015, p. 218.