Structure and Stability Theorems
in Topological Data Analysis

Master's Final Thesis

Author: Gonzalo Ortega Carpintero

Advisor (UPM): Jaime Santos Rodríguez

Advisor (UAM): Luis Guijarro Santamaría

Thursday, July 3, 2025

Universidad Autónoma de Madrid

Contents

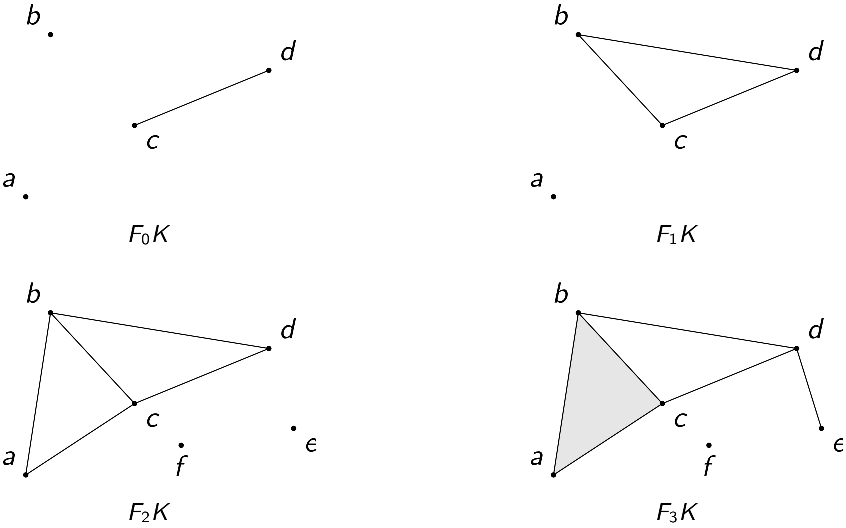


Figure: Four step filtration of a simplicial complex K.

Definition (Persistence module)

Let $\mathbb{F}$ be a field and let T be a totally ordered set. Let $V = \{V_t\}_{t\in T}$ be a collection of F-vector spaces. A T-indexed **persistence module** is a pair (V, π) such that $\pi = \{\pi_{s\leq t}\}$ is a collection of linear maps $\pi_{s\leq t}\colon V_s \to V_t$ that verifies that for all $r, s, t \in T$,

$$\pi_{r\leq s} \circ \pi_{s\leq t} = \pi_{r\leq t}.$$

Definition δ-interleaved modules

Let $(V,\pi),(W,\theta)$ be two persistence modules and let $\delta>0$. V and W are δ-**interleaved** if there exists two persistence module morphisms $\phi\colon V\to W_\delta$ and $\psi\colon W\to V_\delta$ such that the following diagrams commute:

$$\begin{tikzcd} V \arrow[r, "\phi"] \arrow[rr, bend right, "\pi_{2\delta}"'] & W_\delta \arrow[r, "\psi_\delta"] & V_{2\delta} \end{tikzcd}, \qquad \begin{tikzcd} W \arrow[r, "\psi"] \arrow[rr, bend right, "\theta_{2\delta}"'] & V_\delta \arrow[r, "\phi_\delta"] & W_{2\delta} \end{tikzcd}.$$

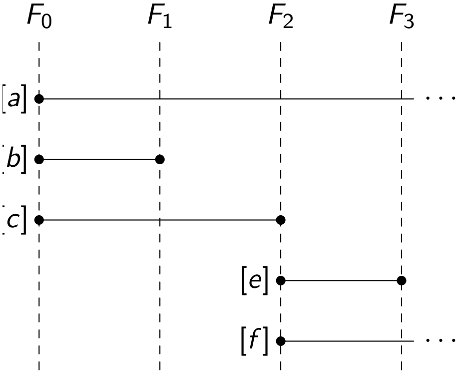
Definition (Interleaving distance)

Let (V, π) and (W, θ) be two tame persistence modules. The **interleaving distance** between them is defined as

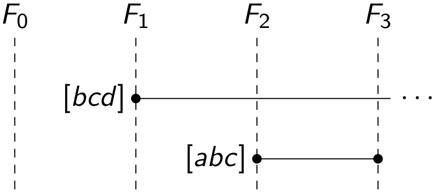
$$d_{\text{int}}(V, W) := \inf\{\delta > 0 \mid V \text{ and } W \text{ are } \delta\text{-interleaved}\}.$$

Definition (Barcode)

A **barcode** B is a finite multiset of intervals. That is, a collection $\{(I_i, m_i)\}$ of intervals I_i with multiplicities $m_i \in N$, where each interval I_i is either finite of the form $(a, b]$ or infinite of the form (a, ∞). Each interval I_i is named to be a **bar** of B. The first number, a is named the **birth** of the barcode and is second number is its **death**.



(a) Bars of $H_0(K)$.



(b) Bars of $H_1(K)$.

Figure: Barcodes associated to the previous filtration.

Definition (Persistence diagram)

Let I be a countable multiset. A *persistence diagram* is a function $D : I \to \mathbb{R}^2_<$.

Definition (Persistence diagram)

Let I be a countable multiset. A *persistence diagram* is a function $D : I \to \mathbb{R}^2_<$.

Definition (Partial mathing)

Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. A *partial matching* between D_1 and D_2 is the triple (I_1', I_2', f) such that $f : I_1' \to I_2'$ is a bijection with $I_1' \subseteq I_1$ and $I_2' \subseteq I_2$.

Definition (p-cost)

Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. Let (I'_1, I'_2, f) be a partial matching between them. If $p < \infty$, the *p-cost of f* is defined as

$$\mathrm{cost}_p(f) := \Big(\sum_{i \in I'_1} d_\infty(D_1(i), D_2(f(i)))^p + \sum_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta)^p + \sum_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta)^p\Big)^{\frac{1}{p}}.$$

For $p = \infty$, the *∞-cost of f* is defined as

$$\mathrm{cost}_\infty(f) := \max\{\sup_{i \in I'_1} d_\infty(D_1(i), D_2(f_i)), \sup_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta), \sup_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta)\}.$$

Definition (Wasserstein distance)

Let D_1, D_2 be persistence diagrams. Let $1 \leq p \leq \infty$. Define

$$\tilde{\omega}_p(D_1, D_2) = \inf\{\mathrm{cost}_p(f) : f \text{ is a partial matching between } D_1 \text{ and } D_2\}.$$

Let $\emptyset$ denote the unique persistence diagram with empty indexing set. Let $(\mathrm{Dgm}_p, \omega_p)$ be the space of persistence diagrams D that satisfy $\tilde{\omega}_p(D, \emptyset) < \infty$ modulo the equivalence relation $D_1 \sim D_2$ if $\tilde{\omega}_p(D_1, D_2) = 0$. The metric ω_p is called the *p-Wasserstein distance*.

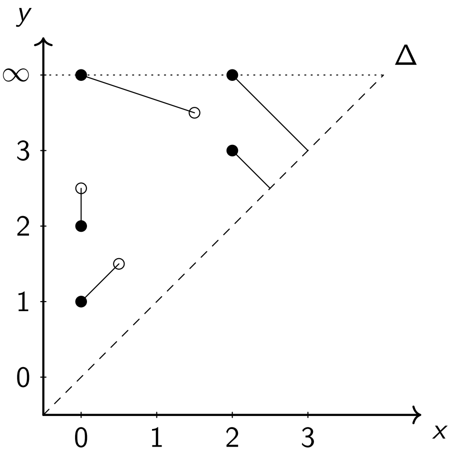


Figure: Wasserstein distance between two persistence diagrams.

Contents

Theorem

Let (V, π) be a persistence module. There exist a barcode $\mathrm{Bar}(V, \pi)$, with $\mu\colon \mathrm{Bar}(V, \pi) \longrightarrow \mathbb{N}$, the multiplicity of the barcode intervals, such that there is a unique direct sum decomposition

$$V \cong \bigoplus_{I \in \mathrm{Bar}(V)} \mathbb{F}(I)^{\mu(I)}.$$

Contents

Theorem

There is an isometry between a persistence module with the interleaving distance and its barcode with the bottleneck distance. That is, given two persistence modules V and W, it holds that

$$d_{\mathrm{int}}(V, W) = d_{\mathrm{bot}}(\mathrm{Bar}(V), \mathrm{Bar}(W)).$$

Theorem

Let X be a triangulable space, and $f, g\colon X \to \mathbb{R}$ continuous tame functions. Then,

$$d_{\mathrm{bot}}(D(f), D(g)) \leq d_{\mathrm{H}}(D(f), D(g)) \leq \|f - g\|_\infty.$$

Theorem

Let (X, d_X), (Y, d_Y) be finite metric spaces. Then, for any $k \in \mathbb{N}$,

$$d_{\text{bot}}(D_k(\mathcal{R}(X, d_X)), D_k(\mathcal{R}(Y, d_Y))) \leq d_{\text{GH}}((X, d_X), (Y, d_Y)).$$

Contents

Definition (Rank function)

The **rank function** of a persistence module V is the function $\delta\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$\lambda(b,d) = \begin{cases} \beta_b^d & \text{if } b \leq d, \\ 0 & \text{otherwise.} \end{cases}$$

Definition (Rank function)

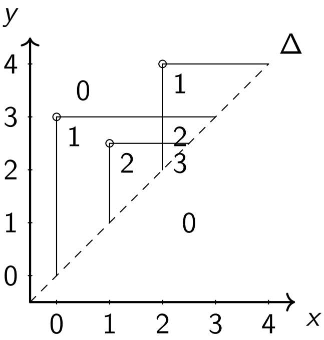
The **rank function** of a persistence module V is the function $\delta\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$\lambda(b,d) = \begin{cases} \beta_b^d & \text{if } b \le d, \\ 0 & \text{otherwise.} \end{cases}$$

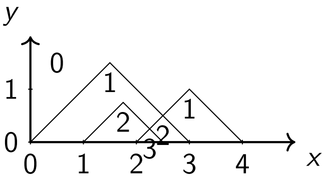
Definition (Persistence landscape)

A **persistence landscape** is a function $\lambda\colon \mathbb{N} \times \mathbb{R} \to \overline{\mathbb{R}}$, defined as

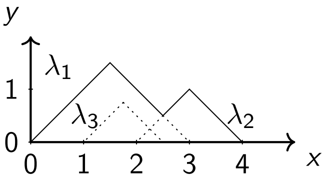
$$\lambda(k,t) \coloneqq \sup\{m \ge 0 \mid \beta^{t-m,t+m} \ge k\}.$$



(a) Rank function.



(b) Rescaled rank function.



(c) Persistence landscape.

Figure: Persistence landscape of a persistence diagram.

Definition (Persistence surface)

The **persistence surface** associated to D, by f and ϕ_u is a function $\rho_D \colon \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\rho_D(z) := \sum_{u \in T(D)} f(u)\phi_u(z).$$

Definition (Persistence surface)

The **persistence surface** associated to D, by f and ϕ_u is a function $\rho_D \colon \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\rho_D(z) := \sum_{u \in T(D)} f(u)\phi_u(z).$$

Definition (Persistence image)

Let D be a persistence diagram with an associated persistence surface ρ_D. The **persistence image** of D by ρ_D is the collection ρ of **pixels**

$$I(\rho_D)_p := \iint_p \rho_B dydx.$$

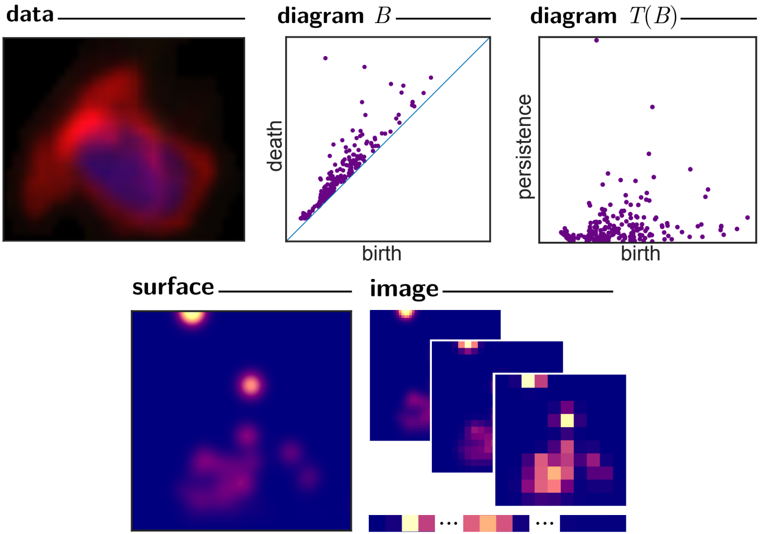


Figure: Algorithm pipeline to transform data into a persistence image.

Definition

Let K be a simplicial complex, and let K^p be its p-skeleton. The **Euler characteristic** of K is the alternating sum of the number of cells in its dimension

$$\chi(K) := \sum_d (-1)^d \#(K^d).$$

Definition

Let K be a simplicial complex, and let K^p be its p-skeleton. The **Euler characteristic** of K is the alternating sum of the number of cells in its dimension

$$\chi(K) := \sum_d (-1)^d \#(K^d).$$

Definition

Let K be a simplicial complex. Let $f\colon K \to \mathbb{R}$ be a filtration function. The **Euler characteristic curve** is a function that assign an Euler characteristic χ for each filtration level $t \in \mathbb{R}$.

$$\mathrm{ECC}(K, t) := \chi(K_t),$$

where $K_t = f^{-1}(-\infty, t]$.