Topological Descriptors for Geometric Data
Reminder: the TDA pipeline

The 5 pillars of the theory (persistence theory):

- decomposition theorems (existence of barcodes)
- algorithms (computation of barcodes)
- stability theorems (barcodes as stable descriptors)
- statistical frameworks for barcodes
- vectorizations and kernels on barcodes for learning
1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power
Outline

1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power
Geometric Data

**Input:** point cloud equipped with a **metric** or **(dis-)similarity measure**

data point ≡ image/patch, geometric shape, protein conformation, patient, LinkedIn user...
Geometric Data

Input: point cloud equipped with a **metric** or **(dis-)similarity measure**

data point $\equiv$ image/patch, geometric shape, protein conformation, patient, LinkedIn user...

Goal: describe the structure of the geometry underlying the data, for interpretation or summary
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

$$d_{GH}(X, Y) := \inf_{s_X : X \to Z, s_Y : Y \to Z} d_Z(s_X(X), s_Y(Y))$$
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

$$d_{\text{GH}} = 0$$
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

$$d_{GH} > 0$$
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

$\text{d}_{\text{GH}} = 0$
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

- descriptor / signature $\equiv$ persistence diagram / feature vector
Why use descriptors

[Agarwal et al. 2015] show that it is NP-hard to approximate the GH distance within a factor of 3, even for metric trees.

Why use descriptors?

- **Shape space**
- **Descriptors space**

- **Isometries**
  - GH distance
  - Hard to compute

- **Equality**
  - Distance
  - Easy to compute

[Agarwal et al. 2015]
Why use descriptors

[Agarwal et al. 2015] show that it is NP-hard to approximate the GH distance within a factor of 3, even for metric trees.

Isometries
GH distance
hard to compute
[Agarwal et al. 2015]

shape space

descriptors space

Ideally, descriptors distance = GH distance

equality
distance
easy to compute
Why use descriptors

Ideally, descriptors distance = GH distance

In reality, descriptors distance ≤ GH distance
Why use descriptors

Some descriptors for images / 3d shapes / metric spaces:

- diameter
- curvature (mean, Gaussian, sectional)
- shape context (distribution of distances)
- heat kernel signature (heat diffusion)
- wave kernel signature (Maxwell’s equations)
- spin image (local neighborhood parametrization)
- SIFT features (local distribution of gradient orientations)
- etc.
Why use descriptors

Some descriptors for images / 3d shapes / metric spaces:

• diameter

• curvature (mean, Gaussian, sectional)

• shape context (distribution of distances)

• heat kernel signature (heat diffusion)

• wave kernel signature (Maxwell’s equations)

• spin image (local neighborhood parametrization)

• SIFT features (local distribution of gradient orientations)

• etc.
Topological descriptors

Input: a finite/compact metric space \((X, d_X)\), a basepoint \(x \in X\)

Construction: a filtration (nested family of sublevel-sets of real-valued function)

Signature: the persistence diagram associated with the filtration
Global topological descriptors

Input: a compact metric space \((X, d_X)\)

Descriptor: \(dgm \mathcal{R}(X, d_X)\) where \(\mathcal{R}\) stands for \textit{Vietoris-Rips filtration}

Barcodes / diagrams

Rips filtration

Offsets filtration
Global topological descriptors

Input: a compact metric space \((X, d_X)\)

Descriptor: \(\text{dgm} \mathcal{R}(X, d_X)\) where \(\mathcal{R}\) stands for Vietoris-Rips filtration

\[
\{x_0, \cdots, x_r\} \in R_t(X, d_X) \iff t \geq \max_{i,j} d_X(x_i, x_j)
\]

\(R_t(X, d_X)\)
Global topological descriptors

Input: a compact metric space \((X, d_X)\)

Descriptor: \(\text{dgm} \mathcal{R}(X, d_X)\) where \(\mathcal{R}\) stands for Vietoris-Rips filtration

\[ \{x_0, \cdots, x_r\} \in R_t(X, d_X) \iff t \geq \max_{i,j} d_X(x_i, x_j) \]
Some examples

geodesic

Euclidean
This equality is fine, since both curves are isometric when equipped with the geodesic distance (their total lengths are the same). The Euclidean distance allows us to differentiate between them.

Some examples
Some examples

This equality is not fine, since the two spaces are not isometric. Note that the equality does not come from the sampling... Fortunately, the Euclidean distance allows us to differentiate between the shapes, even though by a single point.
Theorem: [Chazal, de Silva, O. 2013]
For any compact metric spaces \((X, d_X)\) and \((Y, d_Y)\),
\[ d_{\infty}(dgm \mathcal{R}(X, d_X), \ dgm \mathcal{R}(Y, d_Y)) \leq 2d_{GH}(X, Y). \]
The issue with infinite spaces is that they give rise to infinite Rips complexes, whose filtrations may or may not be well-defined. At least, I will leave the suspense open until Frederic Chazal’s talk.

**Theorem:** [Chazal, de Silva, O. 2013]
For any compact metric spaces \((X, d_X)\) and \((Y, d_Y)\),
\[
d_\infty(\text{dgm} \mathcal{R}(X, d_X), \text{dgm} \mathcal{R}(Y, d_Y)) \leq 2d_{GH}(X, Y).
\]

**Proof outline:**
}\[
(X, d_X) \to (Z, d_Z) \to (\mathbb{R}|X|+|Y|, \ell_\infty)
\]

 finite

\[
\gamma_X \quad \text{id} \quad \gamma_X(X) \sqcup \gamma_Y(Y), d_Z \]

\[
\gamma_Y \quad \text{id} \quad \gamma_Y
\]
Toy application (unsupervised shape classification)

60 shapes (represented as point clouds with approximate geodesic distances)
Toy application (unsupervised shape classification)

computation time $\approx$ 1 hour (pacing phase: bottleneck distances computation)
Toy application (unsupervised shape classification)
Local topological descriptors

Input: a compact metric space \((X, d_X)\), a basepoint \(x \in X\)

Descriptor: \(\text{dgm } d_X(x, \cdot)\)
Some examples
Some examples
Stability

**Theorem: (local descriptors)** [Carrière, O., Ovsjanikov 2015]

Let \((X, d_X)\) and \((Y, d_Y)\) be two compact length spaces with bounded curvature, and let \(x \in X\) and \(y \in Y\). If \(d_{GH}((X, x), (Y, y)) \leq \frac{1}{20} \min\{\rho(X), \rho(Y)\}\), then \(d_\infty(dgm d_X(\cdot, x), dgm d_Y(\cdot, y)) \leq 20 \ d_{GH}((X, x), (Y, y))\).

(adaptation of \(d_{GH}\) to pointed spaces)
Theorem: (local descriptors) [Carrière, O., Ovsjanikov 2015]
Let \((X, d_X)\) and \((Y, d_Y)\) be two compact length spaces with bounded curvature, and let \(x \in X\) and \(y \in Y\). If 
\[
d_{GH}((X, x), (Y, y)) \leq \frac{1}{20} \min\{\varrho(X), \varrho(Y)\},
\]
then 
\[
d_\infty(dgm d_X(\cdot, x), dgm d_Y(\cdot, y)) \leq 20 \ d_{GH}((X, x), (Y, y)).
\]

(adaptation of \(d_{GH}\) to pointed spaces)

(convexity radii)

Prerequisite: 
\[
d_{GH}(X, Y) < \frac{1}{20} \ \min\{\varrho(X), \varrho(Y)\}
\]

\[
X = \bigcirc
\]

\[
Y = [\quad]
\]

\[
d_{GH}(X, Y) < \infty = \varrho(Y)
\]

\[
\forall f, g, \ d_\infty(dgm f, dgm g) = \infty
\]
Toy application (unsupervised shape segmentation)

mapping to $\mathbb{R}^3$ via MDS

$k$-means in $\mathbb{R}^3$
Toy application (unsupervised shape segmentation)

mapping to $\mathbb{R}^3$ via MDS

$k$-means in $\mathbb{R}^3$
Toy application (supervised shape segmentation)

Strategy: use k-NN classifier in diagram space (equipped with $d_\infty$)
Outline

1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power
Pros:
- strong invariance and stability:
  \[ d_\infty(dgm X, dgm Y) \leq \text{cst } d_{GH}(X, Y) \]
- information of a different nature
- flexible and versatile

Cons:
- slow to compare
- space of diagrams is not linear
- positive intrinsic curvature
Persistence diagrams as descriptors for data

A solution: map diagrams to Hilbert space and use kernel trick

$$k(\cdot, \cdot) := \langle \phi(\cdot) | \phi(\cdot) \rangle$$
Kernels for persistence diagrams

State of the Art: define $\phi$ explicitly (vectorization) via:

- images [Adams et al. 2015]
Kernels for persistence diagrams

State of the Art: define $\phi$ explicitly (vectorization) via:

- **images** [Adams et al. 2015]

- **finite metric spaces** [Carrière, O., Ovsjanikov 2015]
Kernels for persistence diagrams

State of the Art: define \( \phi \) explicitly (vectorization) via:

- **images** [Adams et al. 2015]

\[
\begin{bmatrix}
a & b & c \\
b & 4 & 0 & 3 \\
c & 5 & 3 & 0 \\
\end{bmatrix}
\]

- **finite metric spaces** [Carrière, O., Ovsjanikov 2015]

- **polynomial roots or evaluations** [Di Fabio, Ferri 2015] [Kališnik 2016]

\[
\{p_1, \ldots, p_n\} \mapsto (P_1(p_1, \ldots, p_n), \ldots, P_r(p_1, \ldots, p_n), \ldots)
\]
Kernels for persistence diagrams

State of the Art: define $\phi$ explicitly (vectorization) via:

- **images** [Adams et al. 2015]

- **finite metric spaces** [Carrière, O., Ovsjanikov 2015]

- **polynomial roots or evaluations** [Di Fabio, Ferri 2015] [Kališnik 2016]

$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

- **landscapes** [Bubenik 2012] [Bubenik, Dłotko 2015]
Kernels for persistence diagrams

State of the Art: define $\phi$ explicitly (vectorization) via:

- **images** [Adams et al. 2015]

- **finite metric spaces** [Carrière, O., Ovsjanikov 2015]

- **polynomial roots or evaluations** [Di Fabio, Ferri 2015] [Kališnik 2016]

$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

- **landscapes** [Bubenik 2012] [Bubenik, Dłotko 2015]

- **discrete measures**:
  
  $\rightarrow$ histogram [Bendich et al. 2014]

  $\rightarrow$ regularize optimal transport [Carrière, Cuturi, O. 2017]

  $\rightarrow$ convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

  $\rightarrow$ heat diffusion [Reininghaus et al. 2015] + exponential [Kwit et al. 2015]
## Kernels for persistence diagrams

<table>
<thead>
<tr>
<th></th>
<th>Images</th>
<th>Metric Spaces</th>
<th>Polynomials</th>
<th>Landscapes</th>
<th>Discrete Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>ambient Hilbert space</td>
<td>$(\mathbb{R}^d, |\cdot|_2)$</td>
<td>$(\mathbb{R}^d, |\cdot|_2)$</td>
<td>$\ell_2(\mathbb{R})$</td>
<td>$L_2(\mathbb{N} \times \mathbb{R})$</td>
<td>$L_2(\mathbb{R}^2)$</td>
</tr>
<tr>
<td>positive (semi-)definiteness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$|\phi(\cdot) - \phi(\cdot)|_H \leq g(d_P)$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$|\phi(\cdot) - \phi(\cdot)|_H \geq f(d_P)$</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>injectivity</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>universality</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>algorithmic cost</td>
<td>f. map: $O(n^2)$ kernel: $O(d)$</td>
<td>f. map: $O(n^2)$ kernel: $O(d)$</td>
<td>f. map: $O(nd)$ kernel: $O(d)$</td>
<td>$O(n^2)$</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>
# Kernels for persistence diagrams

<table>
<thead>
<tr>
<th></th>
<th>images ($\mathbb{R}^d, |\cdot|_2$)</th>
<th>metric spaces ($\mathbb{R}^d, |\cdot|_2$)</th>
<th>polynomials ($\ell_2(\mathbb{R})$)</th>
<th>landscapes ($L_2(\mathbb{N} \times \mathbb{R})$)</th>
<th>discrete measures ($L_2(\mathbb{R}^2)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ambient Hilbert space</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive (semi-)definiteness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$|\phi(\cdot) - \phi(\cdot)|_H \leq g(d_p)$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$|\phi(\cdot) - \phi(\cdot)|_H \geq f(d_p)$</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>injectivity</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>universality</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>algorithmic cost</td>
<td>f. map: $O(n^2)$, kernel: $O(d)$</td>
<td>f. map: $O(n^2)$, kernel: $O(d)$</td>
<td>f. map: $O(nd)$, kernel: $O(d)$</td>
<td>$O(n^2)$</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>
The idea here is to treat diagrams as measures and to take their densities as feature vectors (to build the feature map, from which the kernel itself is then derived).

Persistence diagrams as discrete measures (I)

\[ \mu_D := \sum_{x \in D} \delta_x \]
Persistence diagrams as discrete measures (I)

\[ \mu_D := \sum_{x \in D} \delta_x \]

\[ \mu^w_D := \sum_{x \in D} w(x) \delta_x \]

**Pb:** \( \mu_D \) is unstable (points on diagonal disappear)

\[ w(x) := \arctan (c \, d(x, \Delta)^r), \ c, r > 0 \]
**Persistence diagrams as discrete measures (I)**

\[ \mu_D := \sum_{x \in D} \delta_x \]

**Pb:** $\mu_D$ is unstable (points on diagonal disappear)

\[ w(x) := \arctan(c \, d(x, \Delta)^r), \quad c, r > 0 \]

**Def:** $\phi(D)$ is the density function of $\mu_D^w \ast \mathcal{N}(0, \sigma)$ w.r.t. Lebesgue measure

\[
\begin{align*}
\phi(D) &:= \frac{1}{\sqrt{2\pi}\sigma} \sum_{x \in D} \arctan(c \, d(x, \Delta)^r) \exp \left( -\frac{\| \cdot - x \|^2}{2\sigma^2} \right) \\
k(D, D') &:= \langle \phi(D), \phi(D') \rangle_{L_2(\Delta \times \mathbb{R}_+)}
\end{align*}
\]
**Persistence diagrams as discrete measures (I)**

- The idea here is to treat diagrams as measures and to take their densities as feature vectors (to build the feature map, from which the kernel itself is then derived).

- **Prop.:** [Kusano, Fukumisu, Hiraoka 2016-17]
  - $\|\phi(D) - \phi(D')\|_H \leq C d_p(D, D')$.
  - $\phi$ is injective and $\exp(k)$ is universal.

\[
\begin{align*}
\phi(D) &:= \frac{1}{\sqrt{2\pi\sigma}} \sum_{x \in D} \arctan(c d(x, \Delta)^r) \exp\left(-\frac{\|\cdot - x\|^2}{2\sigma^2}\right) \\
k(D, D') &:= \langle \phi(D), \phi(D') \rangle_{L_2(\Delta \times \mathbb{R}_+)}
\end{align*}
\]
Metric distortion in practice

![Diagram showing distance in RKHS vs. diagram distance for different kernels: PSS kernel, PWG kernel, SW kernel, and \(\exp(-d_1)\).]
Application to supervised shape segmentation

**Goal**: segment 3d shapes based on examples

**Approach**:
- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape
**Application to supervised shape segmentation**

**Goal**: segment 3d shapes based on examples

**Approach**:  
- train a (multiclass) classifier on PDs extracted from the training shapes  
- apply classifier to PDs extracted from query shape

(training data)
Application to supervised shape segmentation

**Goal**: segment 3d shapes based on examples

**Approach**:
- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape

**Accuracies (%) using TDA descriptors (kernels on barcodes):**

<table>
<thead>
<tr>
<th></th>
<th>TDA</th>
<th>geometry</th>
<th>TDA + geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>74.0</td>
<td>78.7</td>
<td>88.7</td>
</tr>
<tr>
<td>Airplane</td>
<td>72.6</td>
<td>81.3</td>
<td>90.7</td>
</tr>
<tr>
<td>Ant</td>
<td>92.3</td>
<td>90.3</td>
<td>98.5</td>
</tr>
<tr>
<td>FourLeg</td>
<td>73.0</td>
<td>74.4</td>
<td>84.2</td>
</tr>
<tr>
<td>Octopus</td>
<td>85.2</td>
<td>94.5</td>
<td>96.6</td>
</tr>
<tr>
<td>Bird</td>
<td>72.0</td>
<td>75.2</td>
<td>86.5</td>
</tr>
<tr>
<td>Fish</td>
<td>79.6</td>
<td>79.1</td>
<td>92.3</td>
</tr>
</tbody>
</table>
Application to supervised time series analysis

\[ f : \mathbb{N} \rightarrow \mathbb{R} \]

\[ \text{TD}_{m,\tau}(f) := \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+m\tau) \end{bmatrix} \]

\( \tau \): step / delay

\( m\tau \): window size

\( m + 1 \): embedding dimension

Signal

- periodicity
- \# prominent harmonics \((N)\)
- \# non-commensurate freq.

Embedded data

- circularity
- min. ambient dimension \((m \geq 2N)\)
- intrinsic dimension \((S^1 \times \cdots \times S^1)\)

[J. Perea et al.: "SW1PerS: Sliding windows and 1-persistence scoring", 2015]
Application to supervised time series analysis

Contributions of TDA:

- periodicity
- harmonics
- non-commensurate freq.
- underlying state space

no Fourier transform needed

\( f \)
Application to supervised time series analysis

Contributions of TDA:

- inference of:
  - periodicity
  - harmonics
  - non-commensurate freq.
  - underlying state space

no Fourier transform needed

Dynamical system:

**Thm:** [Nash, Takens]
Given a Riemannian manifold $X$ of dimension $m/2$, it is a **generic property** of $\phi \in \text{Diff}_2(X)$ and $\alpha \in C^2(X, \mathbb{R})$ that

$$X \rightarrow \mathbb{R}^{m+1}$$

$$x \mapsto (\alpha(x), \alpha \circ \phi(x), \cdots, \alpha \circ \phi^m(x))$$

is an embedding.
Application to supervised time series analysis

\[
\text{window} \quad \rightarrow \quad \mathbf{f} \quad \rightarrow \quad \text{TD}_{m,\tau} \quad \rightarrow \quad \mathbb{R}^{m+1}
\]

(time-delay embedding)

<table>
<thead>
<tr>
<th>method / dataset</th>
<th>Gyro sensor</th>
<th>EEG dataset</th>
<th>EMG dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM + statistical features</td>
<td>67.6 ± 4.7</td>
<td>44.4 ± 19.8</td>
<td>15.0 ± 10.0</td>
</tr>
<tr>
<td>SVM + Betti sequence</td>
<td>63.5 ± 11.3</td>
<td>66.7 ± 5.6</td>
<td>49.6 ± 18.2</td>
</tr>
<tr>
<td>1-d CNN + dynamic time warping</td>
<td>6.4 ± 5.1</td>
<td>72.4 ± 6.1</td>
<td>15.0 ± 10.0</td>
</tr>
<tr>
<td>imaging CNN</td>
<td>18.9 ± 5.2</td>
<td>48.9 ± 4.2</td>
<td>10.0 ± 0.0</td>
</tr>
<tr>
<td>1-d CNN + Betti sequence</td>
<td>79.8 ± 5.0</td>
<td>75.38 ± 5.7</td>
<td>74.4 ± 10.6</td>
</tr>
</tbody>
</table>

[Y. Umeda: "Time Series Classification via Topological Data Analysis", 2017]
Outline

1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power
Statistics for persistence diagrams

Statistics:
- signal vs noise discrimination
- convergence rates
- confidence indices/intervals, principal components, etc.
Statistics for persistence diagrams

3 approaches for statistics:

- Fréchet means in diagrams space
- embedding into Hilbert spaces
- push-forwards from data space
1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

\[
\text{mean} \left( \begin{array}{c}
\text{bar graph 1} \\
\text{bar graph 2} \\
\text{bar graph 3}
\end{array} \right) = \begin{array}{c}
\text{bar graph result}
\end{array}
\]
1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

\[
\text{mean}(\begin{array}{c}
\text{Diagram 1} \\
\text{Diagram 2} \\
\text{Diagram 3}
\end{array}) = \begin{array}{c}
\text{Diagram}
\end{array}
\]

No coordinates $\mapsto$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \cdots, D_n$:

\[
\bar{D} \in \arg\min_D \frac{1}{n} \sum_i d_p(D, D_i)^2
\]
1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

No coordinates $\rightsquigarrow$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \ldots, D_n$:

$$\bar{D} \in \arg\min_D \frac{1}{n} \sum_i d_p(D, D_i)^2$$

Problem: non-convex energy, highly curved space

$\Rightarrow$ arg min not unique, local minima, numerical issues

1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

\[ \text{mean}(\text{Diagram 1}, \text{Diagram 2}, \text{Diagram 3}) = \text{Diagram 4} \]

No coordinates $\leadsto$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \cdots, D_n$:

\[ \bar{D} \in \arg \min_D \frac{1}{n} \sum_i d_p(D, D_i)^2 \]

Problem: non-convex energy, highly curved space

$\Rightarrow$ arg min not unique, local minima, numerical issues

1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

\[
\text{mean} \left( \begin{array}{c}
\text{-} \\
\text{-} \\
\text{-}
\end{array} \right) \Rightarrow \begin{array}{c}
\text{-} \\
\text{-} \\
\text{-}
\end{array}
\]

New approach: recast problem in measure space

\[ B \mapsto \mu_B \]

\[ \Rightarrow \begin{array}{c}
\text{bitmap} \\
\text{discrete measure}
\end{array} \]

\[ \mu_D := \sum_{x \in D} \delta_x \]

\[ \text{strictly convex problem} \Rightarrow \text{unique mean} \]

easy to compute

\[ \mu_B \mapsto \mu_B \ast \mathcal{U}_{[0, \varepsilon]^2} \]

[M. Agueh, G. Carlier: "Barycenters in the Wasserstein Space", 2011]

\[ W_{2, \gamma}(\mu_{B_i}, \mu_{B_j})^2 := \inf_{\nu} \int \| x - y \|^2 d\nu(x, y) + \gamma H(\nu) \]

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

Rotate PD
Compute rank function

Use boundaries of rank function
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

Use boundaries of rank function

Rotated PD

Compute rank function

\[ x \leq y \implies f^{-1}(-\infty, x) \subseteq f^{-1}(-\infty, y) \]

\[ \nu_x^y : H(f^{-1}(-\infty, x)) \to H(f^{-1}(-\infty, y)) \text{ induced linear map} \]

Rank function is defined as \( \lambda(x, y) = \text{rank} \nu_x^y \)
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

Rotated PD
Compute rank function

Boundaries of rank function: $\lambda_i(t) = \sup\{s \geq 0 : \lambda(t - s, t + s) \geq i\}$

Landscape $\Lambda : \mathbb{R}^2 \rightarrow \mathbb{R}$ is defined as: $\Lambda(i, t) = \lambda_{[i]}(t)$
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

**Prop:** [Bubenik 2015]

\[ \| \Lambda(\text{dgm}) - \Lambda(\text{dgm}') \|_\infty \leq d_\infty(\text{dgm}, \text{dgm}') \]

\( \implies \Lambda \) is Lipschitz hence Borel measurable
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

Prop: [Bubenik 2015]
\[ \| \Lambda(dgm) - \Lambda(dgm') \|_{\infty} \leq d_{\infty}(dgm, dgm') \]

\implies \Lambda \text{ is Lipschitz hence Borel measurable}

Given \( D_1, \cdots, D_n \sim \mu \text{ iid}, \) let \( \overline{\Lambda}^n = \frac{1}{n} \sum_{i=1}^{n} \Lambda(D_i) \)

Thm: (strong law of large numbers) [Bubenik 2015]
If \( E(\| \Lambda(\mu) \|) < +\infty, \) then \( \overline{\Lambda}^n \xrightarrow{a.s.} E(\Lambda(\mu)). \)
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

**Prop:** [Bubenik 2015]
\[ \| \Lambda(dgm) - \Lambda(dgm') \|_{\infty} \leq d_{\infty}(dgm, dgm') \]

\[ \Rightarrow \] \( \Lambda \) is Lipschitz hence Borel measurable

Given \( D_1, \cdots, D_n \sim \mu \) iid, let \( \bar{\Lambda}^n = \frac{1}{n} \sum_{i=1}^{n} \Lambda(D_i) \)

**Thm:** (strong law of large numbers) [Bubenik 2015]
If \( E(\| \Lambda(\mu) \|) < +\infty \), then \( \bar{\Lambda}^n \xrightarrow{a.s.} E(\Lambda(\mu)) \).

**Thm:** (central limit theorem) [Bubenik 2015]
If \( E(\| \Lambda(\mu) \|) < +\infty \) and \( E(\| \Lambda(\mu) \|^2) < +\infty \), then
\[ \sqrt{n} \left( \bar{\Lambda}^n - E(\Lambda(\mu)) \right) \xrightarrow{d} \mathcal{N}(0, \Sigma(\Lambda(\mu))). \]
2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

**Problem:** mean landscape is not a landscape
3. Push-forwards from data space

$$(X, d_X)$$ compact metric space

$\mu$ probability measure supported on $X$ (supp $\mu = X$)

Sample $n$ points iid according to $\mu$.

Examples:
- $\mathcal{F}(\hat{X}_n) = \mathcal{R}(\hat{X}_n, d_X)$
- $\mathcal{F}(\hat{X}_n) = \mathcal{C}(\hat{X}_n, d_X)$
- $\ldots$

Questions:
- Statistical properties of the estimator $dgm \mathcal{F}(\hat{X}_n)$?
- Convergence to the ground truth $dgm \mathcal{F}(X)$? Deviation bounds?
3. Push-forwards from data space

$(X, d_X)$ compact metric space

$\mu$ probability measure supported on $X$ (supp $\mu = X$)

Sample $n$ points iid according to $\mu$.

Examples:
- $\mathcal{F}(\hat{X}_n) = \mathcal{R}(\hat{X}_n, d_X)$
- $\mathcal{F}(\hat{X}_n) = \mathcal{C}(\hat{X}_n, d_X)$
- $\cdots$

Stability thm: $d_\infty(\text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X)) \leq 2d_H(\hat{X}_n, X)$

$\Rightarrow$ for any $\varepsilon > 0$,

$$\mathbb{P} \left( d_\infty \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X), \right) > \varepsilon \right) \leq \mathbb{P} \left( d_H(\hat{X}_n, X) > \frac{\varepsilon}{2} \right)$$
Deviation inequality

For $a, b > 0$, $\mu$ satisfies the $(a, b)$-standard assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

**Theorem** [Chazal, Glisse, Labruère, Michel 2014-15]:
If $\mu$ is $(a, b)$-standard then for any $\varepsilon > 0$:

$$
\mathbb{P} \left( d_\infty \left( \text{dgm} \, \mathcal{F}(\hat{X}_n), \text{dgm} \, \mathcal{F}(X) \right) > \varepsilon \right) \leq \frac{8^b}{a \varepsilon^b} \exp(-n a \varepsilon^b)
$$
Deviation inequality / rate of convergence

For $a, b > 0$, $\mu$ satisfies the $(a, b)$-standard assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

**Theorem** [Chazal, Glisse, Labruère, Michel 2014-15]:
If $\mu$ is $(a, b)$-standard then for any $\varepsilon > 0$:

$$
\mathbb{P} \left( d_{\infty} \left( \text{dgm } F(\hat{X}_n), \text{dgm } F(X) \right) > \varepsilon \right) \leq \frac{8^b}{a \varepsilon^b} \exp \left( -na \varepsilon^b \right)
$$

**Corollary** [Chazal, Glisse, Labruère, Michel 2014-15]:

$$
\sup_{\mu \in \mathcal{P}} \mathbb{E} \left[ d_{\infty} \left( \text{dgm } F(\hat{X}_n), \text{dgm } F(X) \right) \right] \leq C \left( \frac{\log n}{n} \right)^{1/b}
$$

where $C$ depends only on $a, b$. Moreover, the estimator $\text{dgm } F(\hat{X}_n)$ is minimax optimal (up to a $\log n$ factor) on the space $\mathcal{P}$ of $(a, b)$-standard probability measures on $X$. 

27
Numerical illustrations

- $\mu$: unif. measure on Lissajous curve $X$.
- $\mathcal{F}$: distance to $X$ in $\mathbb{R}^2$.
- sample $k = 300$ sets of $n$ points for $n = [2100 : 100 : 3000]$.
- compute
  $$\hat{E}_n = \mathbb{E}[d_\infty(\text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X))].$$
- plot $\log(\hat{E}_n)$ as a function of $\log(\log(n)/n)$.
- $\mu$: unif. measure on a torus $X$.
- $\mathcal{F}$: distance to $X$ in $\mathbb{R}^3$.
- sample $k = 300$ sets of $n$ points for $n = [12000 : 1000 : 21000]$.
- compute
  \[
  \hat{\mathbb{E}}_n = \mathbb{E} \left[ d_\infty \left( \text{dgm} \mathcal{F}(\hat{X}_n), \text{dgm} \mathcal{F}(X) \right) \right].
  \]
- plot $\log(\hat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n \rightarrow \mathcal{F}(\hat{X}_n) \rightarrow \text{dgm } \mathcal{F}(\hat{X}_n)\)

Goal: given \(\alpha \in (0, 1)\), estimate \(c_n(\alpha) \geq 0\) such that

\[
\limsup_{n \to \infty} \mathbb{P} \left( d_\infty \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X) \right) > c_n(\alpha) \right) \leq \alpha
\]

→ confidence region: \(d_\infty\)-ball of radius \(c_n(\alpha)\) around \(\text{dgm } \mathcal{F}(\hat{X}_n)\)
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n \rightarrow \mathcal{F}(\hat{X}_n) \rightarrow \text{dgm } \mathcal{F}(\hat{X}_n)\)

**Goal:** given \(\alpha \in (0, 1)\), estimate \(c_n(\alpha) \geq 0\) such that

\[
\limsup_{n \to \infty} \mathbb{P} \left( d_{\infty} \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X) \right) > c_n(\alpha) \right) \leq \alpha
\]

Note: we already have an inequality of this kind but...

\[
\mathbb{P} \left( d_{\infty} \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X) \right) > \varepsilon \right) \leq \frac{8^b}{\alpha \varepsilon^b} \exp(-n a \varepsilon^b)
\]

unknown
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n \rightarrow F(\hat{X}_n) \rightarrow \text{dgm } F(\hat{X}_n)\)

Goal: given \(\alpha \in (0, 1)\), estimate \(c_n(\alpha) \geq 0\) such that

\[
\limsup_{n \to \infty} \mathbb{P} \left( d_{\infty} \left( \text{dgm } F(\hat{X}_n), \text{dgm } F(X) \right) > c_n(\alpha) \right) \leq \alpha
\]

Bootstrap: (ideally)

- draw \(X^* = X_1^*, \cdots, X_n^*\) iid from \(\mu_{\hat{X}_n}\) (empirical measure on \(\hat{X}_n\))
- compute \(d^* = d_{\infty} \left( \text{dgm } F(X^*), \text{dgm } F(\hat{X}_n) \right)\)
- repeat \(N\) times to get \(d_1^*, \cdots, d_N^*\)
- let \(c_n(\alpha)\) be the \((1 - \alpha)\) quantile of \(\frac{1}{N} \sum_{i=1}^{N} I(d_i^* \geq t)\)

Principle [Efron 1979]: variations of \(\text{dgm } F(X^*)\) around \(\text{dgm } F(\hat{X}_n)\) are same as variations of \(\text{dgm } F(\hat{X}_n)\) around \(\text{dgm } F(X)\).

Note: requires some conditions on \((X, d_X, \mu)\) and diagram space
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n \rightarrow \mathcal{F}(\hat{X}_n) \rightarrow \text{dgm } \mathcal{F}(\hat{X}_n)\)

Goal: given \(\alpha \in (0, 1)\), estimate \(c_n(\alpha) \geq 0\) such that

\[
\limsup_{n \to \infty} \mathbb{P}\left( d_\infty \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X) \right) > c_n(\alpha) \right) \leq \alpha
\]

Bootstrap: (in fact)

- draw \(X^* = X^*_1, \ldots, X^*_n\) iid from \(\mu_{\hat{X}_n}\) (empirical measure on \(\hat{X}_n\))
- compute \(d^* = d_\infty \left( \text{dgm } \mathcal{F}(X^*), \text{dgm } \mathcal{F}(\hat{X}_n) \right) \ d_H(X^*, \hat{X}_n)\)
- repeat \(N\) times to get \(d^*_1, \ldots, d^*_N\)
- let \(c_n(\alpha)\) be the \((1 - \alpha)\) quantile of \(\frac{1}{N} \sum_{i=1}^N I(d^*_i \geq t)\)

Theorem [Balakrishnan et al. 2013] + [Chazal et al. 2014]:

\[
\limsup_{n \to \infty} \mathbb{P}\left( d_\infty \left( \text{dgm } \mathcal{F}(\hat{X}_n), \text{dgm } \mathcal{F}(X) \right) > c_n(\alpha) \right) \leq \alpha.
\]
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n^1, \ldots, \hat{X}_n^m \rightarrow \phi_k(D_n^1), \ldots, \phi_k(D_n^m)\)

\[\bar{v} = \frac{1}{m} \sum_{i=1}^{m} \phi_k(D_n^i)\]

empirical mean feature vector
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n^1, \ldots, \hat{X}_n^m \rightarrow \phi_k(D_n^1), \ldots, \phi_k(D_n^m)\)

\[
\downarrow \\
\text{empirical mean feature vector} \\
\bar{v} = \frac{1}{m} \sum_{i=1}^{m} \phi_k(D_n^i)
\]

Goal: given \(\alpha \in (0, 1)\), estimate \(c_n(\alpha) \geq 0\) such that

\[
\limsup_{n \to \infty} \mathbb{P} \left( \left\| \bar{v} - \mathbb{E}_{\phi_k \circ \text{dgm} \circ \mathcal{F}}^*(\mu^\otimes n)[v] \right\|_{\mathcal{H}_k} > c_n(\alpha) \right) \leq \alpha
\]

mean feature vector according to the measure induced by \(\mu^\otimes n\)

(call it \(\Lambda_{\mu, n}\) for landscapes)
Confidence regions

Setup: \((X, d_X, \mu) \rightarrow \hat{X}_n^1, \ldots, \hat{X}_n^m \rightarrow \Lambda(D_n^1), \ldots, \Lambda(D_n^m)\)

\[
\downarrow
\]

\[
\bar{\Lambda} = \frac{1}{m} \sum_{i=1}^{m} \Lambda(D_n^i)
\]

Bootstrap with landscapes:

- draw \(\Lambda_1^*, \ldots, \Lambda_m^*\) iid from \(\frac{1}{m} \sum_{i=1}^{m} \delta_{\Lambda(D_n^i)}\)
- compute \(\bar{\Lambda}^* = \frac{1}{m} \sum_{i=1}^{m} \Lambda_i^*\) and \(d^* = \|\bar{\Lambda}^* - \bar{\Lambda}\|_{\infty}\)
- repeat \(N\) times to get \(d_1^*, \ldots, d_N^*\)
- let \(c_n(\alpha)\) be the \((1 - \alpha)\) quantile of \(\frac{1}{N} \sum_{i=1}^{N} I(d_i^* \geq t)\)

Theorem [Chazal et al. 2014]:

\[
\limsup_{m \to \infty} \mathbb{P}(\|\bar{\Lambda} - \Lambda_{\mu,n}\|_{\infty} > c_n(\alpha)) \leq \alpha.
\]
Confidence regions

Setup: 

\[(X, d_X, \mu) \rightarrow \hat{X}_n^1, \ldots, \hat{X}_n^m \rightarrow \Lambda(D_n^1), \ldots, \Lambda(D_n^m) \]

\[\downarrow\]

\[\bar{\Lambda} = \frac{1}{m} \sum_{i=1}^{m} \Lambda(D_n^i)\]

Bootstrap with landscapes:

- draw \(\Lambda_1^*, \ldots, \Lambda_m^*\) iid from \(\frac{1}{m} \sum_{i=1}^{m} \delta_{\Lambda(D_n^i)}\)
- compute \(\bar{\Lambda}^* = \frac{1}{m} \sum_{i=1}^{m} \Lambda_i^*\) and \(d^* = ||\bar{\Lambda}^* - \bar{\Lambda}||_\infty\)
- repeat \(N\) times to get \(d_1^*, \ldots, d_N^*\)
- let \(c_n(\alpha)\) be the \((1 - \alpha)\) quantile of \(\frac{1}{N} \sum_{i=1}^{N} I(d_i^* \geq t)\)

**Theorem** [Chazal et al. 2014]:

\[
\limsup_{m \to \infty} \mathbb{P} \left( \left\| \bar{\Lambda} - \Lambda_{\mu,n} \right\|_\infty > c_n(\alpha) \right) \leq \alpha.
\]

\[|\bar{\Lambda}(t) - \Lambda_{\mu,n}(t)|\]

Note: can be done for a fixed \(t\)
Confidence regions

Setup: \( (X, d_X, \mu) \rightarrow \hat{X}_1^n, \ldots, \hat{X}_m^n \rightarrow \Lambda(D_1^n), \ldots, \Lambda(D_m^n) \)

\[
\downarrow \\
\bar{\Lambda} = \frac{1}{m} \sum_{i=1}^m \Lambda(D_i^n)
\]

Bootstrap with landscapes:

- draw \( \Lambda_1^*, \ldots, \Lambda_m^* \) iid from \( \frac{1}{m} \sum_{i=1}^m \delta_{\Lambda(D_i^n)} \)
- compute \( \bar{\Lambda}^* = \frac{1}{m} \sum_{i=1}^m \Lambda_i^* \) and \( d^* = \|\bar{\Lambda}^* - \bar{\Lambda}\|_\infty \)
- repeat \( N \) times to get \( d_1^*, \ldots, d_N^* \)
- let \( c_n(\alpha) \) be the \((1 - \alpha)\) quantile of \( \frac{1}{N} \sum_{i=1}^N I(d_i^* \geq t) \)

**Theorem** [Chazal et al. 2015]:

\[
\|\bar{\Lambda} - \Lambda(\text{dgm } \mathcal{F}(X))\|_\infty \leq \|\bar{\Lambda} - \Lambda_{\mu,n}\|_\infty + \|\Lambda_{\mu,n} - \Lambda(\text{dgm } \mathcal{F}(X))\|_\infty
\]

bias term \( \leq C \left( \frac{\log n}{an} \right)^{1/b} \) when \( \mu \) is \((a, b)\)-standard

variance term
Some applications

**Application 1:** 3D shapes classification

Each mesh has 7K to 40K vertices

From \( m = 100 \) subsamples of size \( n = 300 \)
Some applications

Application 2: walking behaviors classification from smartphone accelerometer data

- spatial time series (accelerometer data from the smartphone of users).
- no registration/calibration preprocessing step needed to compare!
Outline

1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations

![Diagram showing point cloud, offsets filtration, simplicial filtration, and barcode/diagram]
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations

\[ \{x_0, \cdots, x_r\} \in R_t(X, d_X) \iff t \geq \max_{i,j} d_X(x_i, x_j) \]
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations

\[
dgm R(P, \ell_2) = \{(0, +\infty)\} \sqcup \{(0, 1)\} \sqcup \{(0, 1)\}
\]

\[\Rightarrow \text{diagrams for different values of } \alpha \text{ are indistinguishable}\]
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations

**Prop: [Folklore]**
For any *metric tree* \((X, d_X)\):

\[
dgm \mathcal{R}(X, d_X) = \{(0, +\infty)\}
\]

⇒ no information on the metric
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs

⇒ Reeb graphs are indistinguishable from their diagrams
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

**Prop:** [Folklore]

Given \( f : X \to \mathbb{R} \) and \( h : Y \to X \) homeomorphism, \[ \text{dgm} \ f \circ h = \text{dgm} \ f \]

\( \Rightarrow \) Persistence is invariant under reparametrizations
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

Possible solutions:

- Richer topological invariants (e.g. persistent homotopy)
- Use several filter functions (concatenation vs multipersistence)
How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

possible solutions:

- richer topological invariants (e.g. persistent homotopy)
- use several filter functions (**concatenation** vs **multipersistence**)
Persistent Homology Transform (PHT)

\((X, d_X)\) (compact)

\[ \mathcal{F} = \{ f_w \}_{w \in W} \]

\[ \mathbb{R} \]

\[ \text{PHT}(X) = \{ \text{dgm } f_w \mid w \in W \} \]

(diagrams, \(d_\infty\))
Persistent Homology Transform (PHT)

Thm: [Turner, Mukherjee, Boyer 2014]
Let $\mathcal{F} = \{\langle \cdot, w \rangle \}_{w \in S^{d-1}}$, where $d = 2, 3$ is fixed. Then, PHT is injective on the set of linear embeddings of compact simplicial complexes in $\mathbb{R}^d$.

Extension: [Turner et al., in progress]
True for arbitrary $d$ and semialgebraic compact sets.
**Persistent Homology Transform (PHT)**

**Thm:** [Turner, Mukherjee, Boyer 2014]

Let $\mathcal{F} = \{\langle \cdot, w \rangle \}_{w \in \mathbb{S}^{d-1}}$, where $d = 2, 3$ is fixed. Then, PHT is injective on the set of linear embeddings of compact simplicial complexes in $\mathbb{R}^d$.

**Corollary:** PHT is a **sufficient statistic** for such sets $\Rightarrow$ parametric inference
Given $(X, d_X)$ compact length space, take $\mathcal{F} = \{d_X(\cdot, x)\}_{x \in X}$
PHT for intrinsic metrics

Given \((X, d_X)\) compact length space, take \(\mathcal{F} = \{d_X(\cdot, x)\}_{x \in X}\)

**Thm:** [O., Solomon 2017]
There is a Gromov-Hausdorff dense subset of the compact length spaces on which the intrinsic PHT is injective.
Generic injectivity

Generative model:

metric graph \equiv \text{combinatorial graph} (V, E) + \text{edge weights} \ E \to \mathbb{R}_+

mixture (\text{proba. mass function}, \text{proba. measure with density on } \mathbb{R}_+^{\left|E\right|})
Generic injectivity

Generative model:

metric graph $\equiv$ combinatorial graph $(V, E) +$ edge weights $E \rightarrow \mathbb{R}_+$

Thm: [O., Solomon 2017]
Under this model, there is a full-measure subset of the metric graphs on which the intrinsic PHT is injective.
Generic injectivity

Generative model:

metric graph \equiv \text{combinatorial graph} \ (V, E) + \text{edge weights} \ E \rightarrow \mathbb{R}_+

\begin{align*}
\text{mixture (proba. mass function, proba. measure with density on } \mathbb{R}_+^{|E|})
\end{align*}

\textbf{Thm:} [O., Solomon 2017]
Under this model, there is a full-measure subset of the metric graphs on which the intrinsic PHT is injective.

\textbf{Aim:} PHT as a sufficient statistic for metric graphs \Rightarrow \text{parametric inference}
Persistence diagrams as descriptors for data

Pros:
- strong invariance and stability:
  \[ d_\infty(\text{dgm } X, \text{dgm } Y) \leq \text{cst } d_{\text{GH}}(X, Y) \]
- information of a different nature
- flexible and versatile

Pros:
- quasi-isometric maps to Hilbert space
- kernel trick
- provable discriminativility on certain classes of spaces
- (statistics via push-forwards/pull-backs)