Persistent homology and nonparametric regression

Peter Bubenik

Cleveland State University

March 10, 2009, BIRS: Data Analysis using Computational Topology and Geometric Statistics

joint work with Gunnar Carlsson (Stanford), Moo Chung (Wisconsin–Madison), Peter Kim and Zhiming Luo (Guelph)
Introduction

Persistent Homology Statistics Application Framework

Ideal Truth (Parameter)

Observed Reality (Statistic)

Function on a manifold

Sampled Data

Topological Description via Persistent Homology

Filter using level sets

measurement
Introduction Persistent Homology Statistics Application

Framework

Ideal Truth (Parameter)

Function on a manifold

Filter using level sets

Topological Description via Persistent Homology

measurement

Sampled Data

Filtered Simplicial Complex

Topological Description via Persistent Homology

How close?

Observed Reality (Statistic)
Framework

Ideal Truth (Parameter)

Function on a manifold

Filter using level sets

Topological Description via Persistent Homology

measurement

Observed Reality (Statistic)

Sampled Data

Statistical techniques

Estimated Function

Filter using level sets

Topological Description via Persistent Homology

How close?
Persistent homology describes the homological features which persist as a single parameter changes.

Here, we take this parameter to be a threshold on a function on the space from which we are sampling.
Let $f : \mathbb{R} \to \mathbb{R}$. Assume: if $f'(x) = 0$, then $f''(x) \neq 0$, ($f$ has non-degenerate critical points).

Each critical point is a local minimum or a local maximum.
Let \( f : \mathbb{R} \to \mathbb{R} \). Assume: if \( f'(x) = 0 \), then \( f''(x) \neq 0 \), (\( f \) has non-degenerate critical points).

Each critical point is a local minimum or a local maximum.

Define the sublevel sets \( \mathbb{R}_{f \leq t} = f^{-1}(-\infty, t], \ t \in \mathbb{R} \).

**Remark**

As \( t \) increases, the topology of \( \mathbb{R}_{f \leq t} \) does not change as long as we do not pass a critical value.
At the critical points we have the following effects:

- At a local minimum, a new component is added.
- At a local maximum, two components are merged.
At the critical points we have the following effects:

- At a local minimum, a new component is added.
- At a local maximum, two components are merged.

Pairing critical points: Pair a local maximum, with the higher (younger) of the two local minimum associated with the two components which it joins.
At the critical points we have the following effects:

- At a local minimum, a new component is added.
- At a local maximum, two components are merged.

Pairing critical points: Pair a local maximum, with the higher (younger) of the two local minimum associated with the two components which it joins.

If we graph all pairs of critical points we obtain the (Reduced) Persistence Diagram.
For example:

\[ y = f(x) \]
Introduction  Persistent Homology  Statistics  Application  On $\mathbb{R}$  On a manifold  Bottleneck Distance

### Persistence Diagram

For example:

$y = f(x)$

$\mathbb{R}_{f \leq 1}$

Peter Bubenik  Persistent homology and nonparametric regression
For example:

\[ y = f(x) \]

\[ \mathbb{R}_{f \leq 2} \]
Persistent Homology and Nonparametric Regression

For example:

\[ y = f(x), \quad R f \leq 3 \]
For example:

\[ y = f(x) \]

\[ \mathbb{R}_{f \leq 4} \]
For example:

\[ y = f(x) \]

\begin{figure}
\centering
\includegraphics[width=\textwidth]{persistence_diagram}
\caption{Persistence Diagram}
\end{figure}
Let $\mathcal{M}$ be a manifold and let $f : \mathcal{M} \to \mathbb{R}$.

This function gives an increasing filtration of $\mathcal{M}$ by sublevel sets

$$\mathcal{M}_{f \leq r} = \{x \in \mathcal{M} \mid f(x) \leq r\}.$$ 

This induces an increasing filtration on $C_*(\mathcal{M})$. 
Let $\mathcal{M}$ be a manifold and let $f : \mathcal{M} \to \mathbb{R}$.

This function gives an increasing filtration of $\mathcal{M}$ by sublevel sets

$$\mathcal{M}_{f \leq r} = \{ x \in \mathcal{M} \mid f(x) \leq r \}.$$ 

This induces an increasing filtration on $C_*(\mathcal{M})$.

For $s \leq t$, the inclusion $i^t_s : \mathcal{M}_{f \leq s} \to \mathcal{M}_{f \leq t}$ induces

$$H_*(i^b_a) : H_k(\mathcal{M}_{f \leq s}) \to H_k(\mathcal{M}_{f \leq t}),$$

whose image is the persistent homology from $s$ to $t$ of $f$. 
Let $f : \mathcal{M} \to \mathbb{R}$ be a Morse function with distinct critical values $t_0 < t_1 < \cdots < t_k$. Recall that $M_{f \leq t} = f^{-1}(\infty, t]$.

The index $p$ associated to $t_j$ is the number of negative eigenvalues of the Hessian.
Let \( f : \mathcal{M} \rightarrow \mathbb{R} \) be a Morse function with distinct critical values \( t_0 < t_1 < \cdots < t_k \). Recall that \( M_{f \leq t} = f^{-1}(\infty, t] \).

The index \( p \) associated to \( t_j \) is the number of negative eigenvalues of the Hessian.

Morse theory \( \Rightarrow M_{f \leq t_j} \simeq M_{f \leq t_{j-1}} + \) a \( p \)-dimensional cell.
So \( \dim H_p \) increases by one (positive) or \( \dim H_{p-1} \) decreases by one (negative).

Pair: the positive index \( p \) critical values and the index \( p + 1 \) negative critical values \( \Rightarrow \) Persistence Diagram, \( D_p(f) \).
A Function

For example
Sublevel sets and $H_1$
Persistence Diagram of the function
The Bottleneck Distance

A useful metric on the space of Persistence Diagrams:
The Bottleneck Distance

A useful metric on the space of Persistence Diagrams:

Let $f, g : \mathcal{M} \to \mathbb{R}$ be two Morse functions, with associated Persistence Diagrams $D_p(f)$ and $D_p(g)$.

**Definition**

The Bottleneck distance is given by

$$d_B(D_p(f), D_p(g)) = \inf_{\eta} \sup_x \| x - \eta(x) \|_\infty,$$

where the infimum is taken over all bijections $\eta : D_p(f) \to D_p(g)$ and the supremum is taken over all points $x \in D_p(f)$. 

Peter Bubenik

Persistent homology and nonparametric regression
Bottleneck Distance

For example:

\[ y = f(x) \]
Bottleneck Distance

For example:

\[ y = f(x) \]

\[ \begin{array}{cccc}
1 & 2 & 3 & 4 \\
\hline
1 & 2 & 3 & 4
\end{array} \]

\[ \begin{array}{cccc}
birth & & & \\
\hline
1 & 2 & 3 & 4
\end{array} \]

\[ \begin{array}{cccc}
death & & & \\
\hline
1 & 2 & 3 & 4
\end{array} \]
The following fundamental result bounds the bottleneck distance for persistence diagrams with the supremum norm.

**Theorem (Cohen-Steiner, Edelsbrunner, Harer)**

\[ d_B(D_p(f), D_p(g)) \leq \| f - g \|_\infty \]
The following fundamental result bounds the bottleneck distance for persistence diagrams with the supremum norm.

**Theorem (Cohen-Steiner, Edelsbrunner, Harer)**

\[ d_B(D_p(f), D_p(g)) \leq \|f - g\|_{\infty} \]

For us, this result is crucial as it allows us to connect topology to statistics.
The statistical viewpoint

Assume that $f = f_\theta$ belongs to a family of functions

$$\{f_\theta \mid \theta \in \Theta\}$$

where $\theta$ is the parameter and $\Theta$ is the parameter space which can be finite dimensional (parametric), or, infinite dimensional (nonparametric).

Goal

Find an estimate $\hat{\theta}$ of $\theta$ so that the persistent homology of $f_\hat{\theta}$ is close to the persistent homology of $f_\theta$. 
Stability and statistics

Take $f$ to be an unknown function and $\hat{f}$ to its statistical estimator. Then,

$$d_B(D_p(f), D_p(\hat{f})) \leq \|f - \hat{f}\|_{\infty}$$
Let $\mathcal{M}$ be a manifold.

**Problem (Nonparametric regression problem)**

Assume that there exists a function $f : \mathcal{M} \rightarrow \mathbb{R}$ such that

$$y = f(x) + \epsilon, \quad x \in \mathcal{M}$$

where $\epsilon$ is a normal random variable with mean zero and variance $\sigma^2 > 0$. Given a sample $(x_1, y_1), \ldots, (x_n, y_n)$, find an estimator $\hat{f}$ of $f$.

Goal: find an estimator $\hat{f}$ that minimizes $\|f - \hat{f}\|_\infty$ and calculate the asymptotics as $n \rightarrow \infty$. 

Peter Bubenik

Persistent homology and nonparametric regression
\[ \mathcal{M} \text{ is a compact } d\text{-dimensional Riemannian manifold with metric } \rho(\cdot, \cdot) \text{ (given by geodesic distance).} \]
The precise setup

\[ M \] is a compact \( d \)-dimensional Riemannian manifold with metric \( \rho(\cdot, \cdot) \) (given by geodesic distance).

Parameter space: the Hölder class of functions,

\[
\Lambda(\beta, L) = \left\{ f : M \to \mathbb{R} : |f(x) - f(y)| \leq L \rho(x, y)^\beta, \ x, y \in M \right\},
\]

where \( 0 < \beta \leq 1 \).
Expected loss and minimax risk

Definition

- The expected loss (risk) of an estimator \( \tilde{f} \) is given by

\[
E\|\tilde{f} - f\|_\infty,
\]

where \( E \) is with respect to \( y = f(x) + \varepsilon \).

- The minimax risk is given by

\[
\inf_{\tilde{f}} \sup_{f \in \Lambda(\beta,L)} E\|\tilde{f} - f\|_\infty.
\]
The main result

**Theorem (B-C-C-K-L)**

There exists an estimator $\hat{f}$ (constructive) that attains the minimax risk, and

$$\sup_{f \in \Lambda(\beta,L)} \mathbb{E} \| \hat{f} - f \|_\infty \sim C \psi_n, \quad \text{as } n \to \infty,$$

where $\psi_n = \left( \frac{\log(n)}{n} \right)^{\beta/(2\beta+d)}$ and

$$C = L^{d/(2\beta+d)} \left( \frac{\sigma^2 \text{vol}(M)(\beta+d)d^2}{\text{vol}(S^{d-1})\beta^2} \right)^{\beta/(2\beta+d)}.$$
The construction of the estimator

A sketch of the construction of the estimator:

Partition $\mathcal{M}$ by $\{A_i \subset \mathcal{M}\}$ and let

$$\hat{f}(x) = \sum_i \hat{a}_i l_{A_i}(x).$$

By a suitable choice of $A_i, \hat{a}_i$, (constructive) one obtains the desired estimator.
Corollary

In the regression model,

\[ \mathbb{E}d_B(D_p(\hat{f}), D_p(f)) \leq C \psi_n \]

as \( n \to \infty \).
Application to Brain Imaging

http://brainimaging.waisman.wisc.edu

Waisman Laboratory for Brain Imaging and Behavior
3T MRI, PET, microPET, EEG, MEG, eye tracking, etc.
everything under a single roof. Research only facility.
6 faculty + 10 PhD staff scientists + 100 students/postdocs
3 Tesla Magnetic Resonance Imaging

1 brain image
= 200 x 200 x 100 array
= 4 million measurements

16 autistic &
12 normal controls
age matched right-handed males
MRI Neuroanatomy

Cerebral Spinal Fluid (CSF)

- Outer Cortical Surface
- Gray Matter
- Inner Cortical Surface
- White Matter

MRI

Peter Bubenik

Persistent homology and nonparametric regression
Cortical surface
Cortex thickness
Cortex thickness
Constructing our estimator

Construct an estimator:
First, smooth the data on $S^2$ using the kernel

$$K_{x_i}(x) = \max(1 - \kappa \arccos(x_i'x), 0),$$

and the usual kernel function estimator

$$\tilde{f}(x) = \frac{\sum_i y_i K_{x_i}(x)}{\sum_i K_{x_i}(x)}. \quad (1)$$
Construct an estimator:
First, smooth the data on $S^2$ using the kernel

$$K_{x_i}(x) = \max(1 - \kappa \arccos(x'_i x), 0),$$

and the usual kernel function estimator

$$\tilde{f}(x) = \frac{\sum_i y_i K_{x_i}(x)}{\sum_i K_{x_i}(x)}.$$ (1)

Next, choose design points from a triangulation of the sphere: take an iterated subdivision of the icosahedron, which has 1280 faces and 642 vertices.
Constructing our estimator:

First, smooth the data on $S^2$ using the kernel

$$K_{x_i}(x) = \max(1 - \kappa \arccos(x_i'x), 0),$$

and the usual kernel function estimator

$$\tilde{f}(x) = \frac{\sum_i y_i K_{x_i}(x)}{\sum_i K_{x_i}(x)}. \quad (1)$$

Next, choose design points from a triangulation of the sphere: take an iterated subdivision of the icosahedron, which has 1280 faces and 642 vertices.

Define $\hat{f}$ on vertices using (1) and extend by affine interpolation.
Triangulated sphere
Cortex thickness estimator

The estimator
Calculating Persistent Homology

Remark

- Critical points only occur at vertices.
- The values of the estimator at the vertices, induce a filtration of the triangulation of the sphere.
- The persistent homology of this filtered complex is identical to the persistent homology of the estimator.

Use Plex to calculate the persistent homology of the filtered complex.
Persistence diagrams
Cumulative Persistence diagrams
Cumulative Persistence diagrams

New inference and classification under development.
Summary

- Given a function on manifold one can study the persistent homology of its sublevel sets.
- Estimating such a function from data is a nonparametric regression problem.
- The sup–norm minimax estimator gives an estimate for persistent homology.
- This method shows promise for classifying large data sets.