Point-cloud topology
Data sampled from an unknown topological space $Y$. Estimate Betti numbers of $Y$ from the sample.
The standard pipeline

hidden/unknown space \( X \)

finite sample \( Y \subset X \)

simplicial complex \( S = S(Y) \)

homology invariants of \( S \)

\( b_0 = 1 \)
\( b_1 = 1 \)
\( b_2 = 0 \)
The standard pipeline

- hidden/unknown space $X$
- finite sample $Y \subset X$
- simplicial complex $S = S(Y)$
- homology invariants of $S$

$\begin{align*}
b_0 &= 1 \\
b_1 &= 1 \\
b_2 &= 0
\end{align*}$
Simplicial reconstructions

- Given a collection of points $X$ in Euclidean space:
  - Proximity graph
    $$\{\text{all vertices } [x]\} \cup \{\text{edges } [x, y] \text{ such that } \|x - y\| \leq r\}$$
  - Vietoris–Rips complex
    $$\{\text{simplices } [x_0, x_1, \ldots, x_k] \text{ for which every } \|x_i - x_j\| \leq r\}$$
  - Čech complex
    $$\{\text{simplices } [x_0, x_1, \ldots, x_k] \text{ whose vertices are contained in an } (r/2)\text{-ball}\}$$
  - Alpha shape (Edelsbrunner, Kirkpatrick, Seidel 1983)
    $$\left\{\begin{array}{l}
    \text{simplices } [x_0, x_1, \ldots, x_k] \text{ whose vertices are contained in an } (r/2)\text{-ball whose interior meets no other points of } X
    \end{array}\right\}$$

- Desire theorems of the form:
  - If $Y$ is well-sampled from $X$ then $S(Y) \approx X$

Proximity Graph

(picture credit: Elizabeth Meckes)
Vietoris–Rips complex
Čech complex

(picture credit: Elizabeth Meckes)
Properties

- Each complex depends on a scale parameter \( r \)
- \( r=0 \)
  - discrete collection of vertices
- \( r=\infty \)
  - graph complex = the complete graph on \( X \)
  - Vietoris–Rips = the complete simplex on \( X \)
  - Čech = the complete simplex on \( X \)
  - Alpha = Delaunay triangulation of convex hull of \( X \)
- Seek interesting topology in the range \( 0 < r < \infty \)
Persistence
Instability

- Betti numbers are discrete

- Topological spaces
  - topological spaces are continuous
  - the space of topological spaces is discrete

- Finite point-clouds
  - point-clouds are discrete
  - the space of point-clouds is continuous

- Therefore, raw Betti numbers are
  - ✔ suitable for topological spaces
  - ✗ dangerous for point-clouds
Instability

- Betti numbers are discrete

- Topological spaces
  - topological spaces are continuous
  - the space of topological spaces is discrete

- Finite point-clouds
  - point-clouds are discrete
  - the space of point-clouds is continuous

- Therefore, raw Betti numbers are
  - ✔ suitable for topological spaces
  - ✘ dangerous for point-clouds
Instability

- Betti numbers are discrete

- Topological spaces
  - topological spaces are continuous
  - the space of topological spaces is discrete

- Finite point-clouds
  - point-clouds are discrete
  - the space of point-clouds is continuous

- Therefore, raw Betti numbers are
  - ✔ suitable for topological spaces
  - ❌ dangerous for point-clouds
The standard pipeline (first attempt)

hidden/unknown space $Y$

finite sample $X \subset Y$

simplicial complex $S = S(X)$

homology invariants of $S$

$b_0 = 1$
$b_1 = 1$
$b_2 = 0$
The standard pipeline

(hidden/unknown space $Y$)

(finite sample $X \subset Y$)

(filtered complex $S(r) = S(X, r)$)

(persistent homology of $S(r)$)
Persistent homology

- Homology provides functors \( H = H_k \)
- Construct a sequence of spaces
  \[ X_0 \to X_1 \to \cdots \to X_k \]
- Obtain a sequence of vector spaces
  \[ H(X_0) \to H(X_1) \to \cdots \to H(X_k) \]
- Describe the structure of such a sequence
  (what are the irreducible factors?)

one vector space ↔ dimension

sequence of vector spaces ↔ persistence barcode

Edelsbrunner, Letscher, Zomorodian (2000)
Zomorodian, Carlsson (2002)
Cohen-Steiner, Edelsbrunner, Harer (2007)
Persistence

- Algorithm (Edelsbrunner, Letscher, Zomorodian '00)
  - barcode: finite collection of half-open intervals
  - \([b,d)\) indicates feature lifetime: born at time \(b\), dies at time \(d\)

- Stability theorem (Cohen-Steiner, Edelsbrunner, Harer '07)
  - barcode depends continuously on the underlying data
  - interleaved systems have similar barcode (Chazal, Cohen-Steiner, Glisse, Guibas, Oudot '09)
  - continuous measurements (interval length)
  & discrete information (number of intervals)
Example

b_0: (clusters)

b_1: (holes)
Example

$\mathbf{b}_0$: (clusters)

$\mathbf{b}_1$: (holes)
Example

$b_0$: (clusters)

$b_1$: (holes)
Example

\[ b_0: \text{(clusters)} \]

\[ b_1: \text{(holes)} \]

8

0
Example

\( b_0: \) (clusters)

\( b_1: \) (holes)
Example

b₀: (clusters)

b₁: (holes)
Example

\[ b_0 : \text{(clusters)} \]

\[ b_1 : \text{(holes)} \]
Example

$\mathbf{b}_0$: (clusters)

$\mathbf{b}_1$: (holes)
Example

b₀: (clusters)

b₁: (holes)
Example

b₀: (clusters)

b₁: (holes)
Example

\[ b_0: \text{(clusters)} \]

\[ 1 \]

\[ b_1: \text{(holes)} \]

\[ 1 \]
Example

\[ b_0: \text{(clusters)} \]

\[ \begin{array}{c}
1 \\
1
\end{array} \]

\[ b_1: \text{(holes)} \]

\[ \begin{array}{c}
\text{---} \\
\text{---} \\
\text{---} \\
\text{---} \\
\end{array} \]
Example

<table>
<thead>
<tr>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>b₀: (clusters)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>b₁: (holes)</td>
</tr>
</tbody>
</table>
Example

b₀: (clusters)

1

b₁: (holes)

0
Example

\[b_0: \text{clusters}\]

\[b_1: \text{holes}\]
Example

\( b_0: \text{(clusters)} \)

\( b_1: \text{(holes)} \)
Persistence diagram
Persistence diagram

Barcode
Persistence diagram

Persistence diagram

Barcode

points (b,d)

intervals [b,d)
Persistence diagram

Persistence diagram

Barcode

points (b,d) ↔ intervals [b,d]
And the Oscar goes to...

Witness Complexes -- Mumford Dataset
Vin de Silva & Gunnar Carlsson
Visual Image Patches
Visual image patches

- Lee, Pedersen, Mumford (2003) studied the local statistical properties of natural images (from Van Hateren’s database)

- 3-by-3 pixel patches with high contrast between pixels: are some patches more likely than others?

- Carlsson, VdS, Ishkhanov, Zomorodian (2004/8): topological properties of high-density regions in pixel-patch space
The space of image patches

- ~4.2 million high-contrast 3-by-3 patches selected randomly from images in database.

- Normalise each patch twice: subtract mean intensity, then rescale to unit norm.

- Normalised patches live on a unit 7-sphere in 8-dimensional space with the following basis:
High-density regions

- LPM2003 found that the distribution of patches is dense in the 7-sphere.

- There are high-density regions:
  - edge features

- Can we describe the structure of the high-density regions?
  - threshold by k-nearest-neighbour density estimator
Defining “high-density”

- How do we define “high density”?
  - Select a positive integer $k$.
  - $r_k(x) =$ distance between $x$ and its $k$-th nearest neighbour.
  - $x$ is a high-density point $\iff r_k(x)$ is small.

- Select “cuts” by thresholding on $r_k(x)$.
  - $k$ small $\iff$ fine structure
  - $k$ large $\iff$ coarse structure
### Straining a data soup

<table>
<thead>
<tr>
<th>$K_{test}$</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>100%</th>
</tr>
</thead>
</table>
Varying the density parameter

(toy example)
Varying the density parameter

(toy example)
A small platter of cuts

<table>
<thead>
<tr>
<th>K</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=15</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>K=100</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>K=300</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>
### A small platter of cuts

<table>
<thead>
<tr>
<th></th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K=100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K=300</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table contains placeholder text and cannot be accurately rendered.
8-dimensional data
The primary circle

- The thick $e_1$–$e_2$ circle consists of linear gradient patches and their nearby edge feature patches.
8-dimensional data
3-circles model
3-circles model
3 circles explained

linear gradients

vertical features

horizontal features
Why is there a predominance of **vertical/horizontal** local features?

Artefact of the square patch shape?
Artefact of the natural world?
Tilting the camera

orthogonal images

diagonal images
Tilting the camera

orthogonal images

diagonal images
Tilting the camera

orthogonal images

diagonal images
Homogenizing over all tilt angles

- $e_1$-$e_2$ circle: arbitrary linear functions $ax+by$ in the image plane.
- $e_1$-$e_3$ circle: quadratic functions of $x$.
- $e_3$-$e_4$ circle: quadratic functions of $y$.
- What about quadratic functions of arbitrary linear functions $ax+by$?
The Klein bottle

Figure 9. Klein bottle embedded naturally in the parameter space as a completion of the 3-circle model. In the unfolded identification space shown, the primary circle wraps around the horizontal axis twice. The two secondary circles each wrap around the vertical axis once (note: the circle on the extreme left and right are glued together with opposite orientation). Each secondary circle intersects the primary circle twice.


