Persistent Homology II

Michal Hrbek

Charles University - Faculty of Mathematics and Physics

November 24, 2016

Outline: (note the rapidly decreasing amount of rigour)

1. Persistent modules and Carlsson's decomposition theorem (some real mathematics),
2. Methodology of using persistent homology to discover topological features of point clouds (a legit method of data mining),
3. The experiment of Carlsson et al. - topological features of the space of 3x3 patches of natural images (some proper shamanism).

Persistent modules

Let k be a field and $(P, <)$ a totally ordered set with a dense countable subset (in our applications, typically either $\mathbb{N}$ or $\mathbb{R}$ with the usual order).

Let k be a field and $(P, <)$ a totally ordered set with a dense countable subset (in our applications, typically either $\mathbb{N}$ or $\mathbb{R}$ with the usual order).
A *(P-)persistent module M* is a funtor from P to the category of k-vector spaces. (Here, poset P is viewed naturally as a category.)

Persistent modules

Let k be a field and $(P, <)$ a totally ordered set with a dense countable subset (in our applications, typically either $\mathbb{N}$ or $\mathbb{R}$ with the usual order).

A *(P-)persistent module* M is a funtor from P to the category of k-vector spaces. (Here, poset P is viewed naturally as a category.)

Explicitly: Data for M consist of vector spaces $V_p, p \in P$ and k-linear maps $f_{pq}^M : V_p \to V_q$ for all $p \leq q$, together with conditions:

- $f_{pp}^M = \mathrm{id}_{V_p}$ for all $p \in P$,
- $f_{qr}^M \circ f_{pq} = f_{pq}$ for all $p \leq q \leq r \in P$.

Interval modules

A subset I of $(P, <)$ is an interval if $p \leq r \leq q \in P$, where $p, q \in I$, implies that $r \in I$.

Interval modules

A subset I of $(P, <)$ is an interval if $p \leq r \leq q \in P$, where $p, q \in I$, implies that $r \in I$.

We can assign a P-persistent module M_I to I in the following way:

- $M_I(p) = \begin{cases} k, p \in I \\ 0, p \notin I \end{cases}$,
- $M_I(p < q) = \begin{cases} \mathrm{Id}_{\mathrm{k}}, p, q \in I \\ 0, \text{othewrise} \notin I \end{cases}$

A P-persistent module M is pointwise finite, if $M(p)$ is finite dimensional for each $p \in P$.

Carlsson's Theorem

A P-persistent module M is pointwise finite, if $M(p)$ is finite dimensional for each $p \in P$.

Theorem (Carlsson, Crawley-Boevey)

Under certain mild conditions for finiteness of M (satisfied e.g. when M is pointwise finite), any P-persistent module M is isomorphic direct sum of interval modules.

A P-persistent module M is pointwise finite, if $M(p)$ is finite dimensional for each $p \in P$.

Theorem (Carlsson, Crawley-Boevey)

Under certain mild conditions for finiteness of M (satisfied e.g. when M is pointwise finite), any P-persistent module M is isomorphic direct sum of interval modules.

What does isomorphic mean?

Let M, N be P-persistent modules. A homomorphism $f : M \to N$ of P-persistent modules is a collection of k-linear maps $\varphi_p : M(p) \to N(p)$ such that the following diagram commutes for any $p < q \in P$:

$$\begin{array}{ccc} M(p) & \xrightarrow{f^M_{pq}} & M(q) \\ \varphi_p \downarrow & & \varphi_q \downarrow \\ N(p) & \xrightarrow{f^N_{pq}} & N(q) \end{array}$$

Let M, N be P-persistent modules. A homomorphism $f : M \to N$ of P-persistent modules is a collection of k-linear maps $\varphi_p : M(p) \to N(p)$ such that the following diagram commutes for any $p < q \in P$:

$$\begin{array}{ccc} M(p) & \xrightarrow{f^M_{pq}} & M(q) \\ \varphi_p \downarrow & & \varphi_q \downarrow \\ N(p) & \xrightarrow{f^N_{pq}} & N(q) \end{array}$$

In fancy words, $\varphi : M \to N$ is a natural transformation of functors N and M.

Let M, N be P-persistent modules. A homomorphism $f : M \to N$ of P-persistent modules is a collection of k-linear maps $\varphi_p : M(p) \to N(p)$ such that the following diagram commutes for any $p < q \in P$:

$$\begin{tikzcd} M(p) \arrow[r, "f^M_{pq}"] \arrow[d, "\varphi_p"'] & M(q) \arrow[d, "\varphi_q"'] \\ N(p) \arrow[r, "f^N_{pq}"] & N(q) \end{tikzcd}$$

In fancy words, $\varphi : M \to N$ is a natural transformation of functors N and M.
We say that φ is an isomorphism of persistent modules if $\varphi(p)$ is an isomorphism for each $p \in P$.

Carlsson's Theorem

A P-persistent module M is pointwise finite, if $M(p)$ is finite dimensional for each $p \in P$.

Theorem (Carlsson, Crawley-Boevey)

Under certain mild conditions for finiteness of M (satisfied e.g. when M is pointwise finite), any P-persistent module M is isomorphic direct sum of interval modules.

What is a direct sum?

Let $M_i, i \in I$ be a collection of P-peristent modules. Then we define the direct sum $\bigoplus_{i \in I} M_i$ to be the P-persistent module defined as follows:

- $\bigoplus_{i \in I} M_i(p) = \bigoplus_{i \in I} M_i(p)$,
- $\bigoplus_{i \in I} M_i(p < q) = \bigoplus_{i \in I} M_i(p < q)$.

Let X be a point cloud - a finite subset of $\mathbb{R}^d$.

Methodology

Let X be a point cloud - a finite subset of $\mathbb{R}^d$.

- Choose a method of turning X into a simplicial complex - Čech complex, Vietoris-Rips, witness...

Methodology

Let X be a point cloud - a finite subset of $\mathbb{R}^d$.

- Choose a method of turning X into a simplicial complex - Čech complex, Vietoris-Rips, witness...
- Compute the simplicial complex $C_\epsilon(X)$ for each parameter $\epsilon > 0$ (well, not really of course)

Methodology

Let X be a point cloud - a finite subset of $\mathbb{R}^d$.

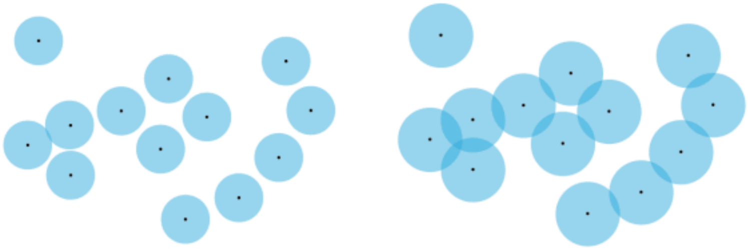
- Choose a method of turning X into a simplicial complex - Čech complex, Vietoris-Rips, witness...
- Compute the simplicial complex $C_\epsilon(X)$ for each parameter $\epsilon > 0$ (well, not really of course)
- X is finite, so there is only finitely many steps, in which the complex changes.

Methodology

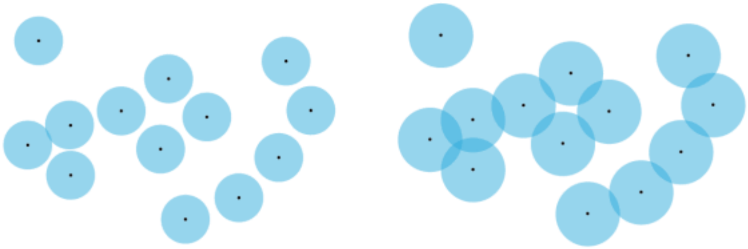
Let X be a point cloud - a finite subset of $\mathbb{R}^d$.

- Choose a method of turning X into a simplicial complex - Čech complex, Vietoris-Rips, witness...
- Compute the simplicial complex $C_\epsilon(X)$ for each parameter $\epsilon > 0$ (well, not really of course)
- X is finite, so there is only finitely many steps, in which the complex changes.
- Let $0 = \epsilon_0 < \epsilon_1 < \cdots < \epsilon_n < \epsilon_{n+1} = \infty$ be parameters such that $C_{\epsilon_i}(X) = C_\delta(X)$ for any $\epsilon_i < \delta < \epsilon_{i+1}$ and $i < n+1$.

$\epsilon_i < \epsilon_j$



$\epsilon_i < \epsilon_j$



$C_{\epsilon_i}(X) \hookrightarrow C_{\epsilon_j}(X)$

- Rename $C_i = C_{\epsilon_i}(X)$.

- Rename $C_i = C_{\epsilon_i}(X)$.

- Rename $C_i = C_{\epsilon_i}(X)$.
- We obtain an increasing sequence of complexes

$$C_0 \hookrightarrow C_1 \hookrightarrow \cdots \hookrightarrow C_{n+1}.$$

- Rename $C_i = C_{\epsilon_i}(X)$.
- We obtain an increasing sequence of complexes

$$C_0 \hookrightarrow C_1 \hookrightarrow \cdots \hookrightarrow C_{n+1}.$$

- By functoriality, this filtration yields maps on k-th homologies for each $k \geq 0$:

$$H_k(C_0) \to H_k(C_1) \to \cdots \to H_k(C_{n+1}).$$

- Rename $C_i = C_{\epsilon_i}(X)$.
- We obtain an increasing sequence of complexes

$$C_0 \hookrightarrow C_1 \hookrightarrow \cdots \hookrightarrow C_{n+1}.$$

- By functoriality, this filtration yields maps on k-th homologies for each $k \geq 0$:

$$H_k(C_0) \to H_k(C_1) \to \cdots \to H_k(C_{n+1}).$$

- Thus, we get for each $k \geq 0$ an $\mathbb{N}$-persistent module H_k.

- By Carlsson's Theorem, H_k decomposes into a direct sum of interval modules $\bigoplus_{i=1}^{l} M(I_i)$.

- By Carlsson's Theorem, H_k decomposes into a direct sum of interval modules $\bigoplus_{i=1}^{l} M(I_i)$.
- The intervals I_i, a priori intervals of $\mathbb{N}$, are interpreted as intervals of $\mathbb{R}$ with borders from $\{\epsilon_0, \epsilon_1, \ldots, \epsilon_{n+1}\}$.

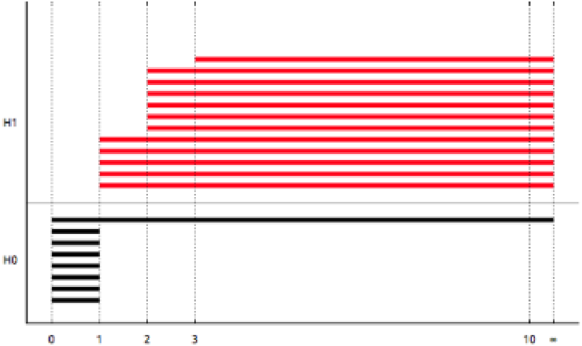
- By Carlsson's Theorem, H_k decomposes into a direct sum of interval modules $\bigoplus_{i=1}^{l} M(I_i)$.
- The intervals I_i, a priori intervals of $\mathbb{N}$, are interpreted as intervals of $\mathbb{R}$ with borders from $\{\epsilon_0, \epsilon_1, \ldots, \epsilon_{n+1}\}$.
- Each interval $I_i = (\epsilon_i, \epsilon_j)$ corresponds to a "birth" of a non-trivial homology cycle (think of a "hole") at parameter ϵ_i, and its "death" at parameter ϵ_j.

Barcodes

Complete topological information obtained by this method in each homological dimension is given by a multiset of intervals. We call this multiset a barcode.

Barcodes

Complete topological information obtained by this method in each homological dimension is given by a multiset of intervals. We call this multiset a barcode.



3x3 patches of natural images

In what follows, "they"=Carlsson et al.:

3x3 patches of natural images

In what follows, “they”=Carlsson et al.:

- Input data: Large collection of grayscale natural images.

3x3 patches of natural images

In what follows, "they"=Carlsson et al.:

- Input data: Large collection of grayscale natural images.
- From each image, they choose 5000 random 3x3-pixel patches. This gives a point cloud in $\mathbb{R}^9$.

3x3 patches of natural images

In what follows, “they”$=$Carlsson et al.:

- Input data: Large collection of grayscale natural images.
- From each image, they choose 5000 random 3x3-pixel patches. This gives a point cloud in $\mathbb{R}^9$.
- Compute a “D-norm” of each patch measuring its contrast (certain quadratic form of logs of intensities of pixels).

3x3 patches of natural images

In what follows, "they"=Carlsson et al.:

- Input data: Large collection of grayscale natural images.
- From each image, they choose 5000 random 3x3-pixel patches. This gives a point cloud in $\mathbb{R}^9$.
- Compute a "D-norm" of each patch measuring its contrast (certain quadratic form of logs of intensities of pixels).
- Keep for each image only those patches being in the top 20% with respect to D-norm.

- Mean center the data - substract a mean of intensity from each pixel. That is, two patches are now considered identical if their “brightness” differ by a constant. This puts all the patches into a 8-dimensional subspace of $\mathbb{R}^9$.

- Mean center the data - substract a mean of intensity from each pixel. That is, two patches are now considered identical if their "brightness" differ by a constant. This puts all the patches into a 8-dimensional subspace of $\mathbb{R}^9$.
- Normalize with respect to the D-norm (normalize the contrast). That is, two patches are now considered identical if they only differ by "turning up contrast". This puts the data on a 7-dimensional ellipsoid inside $\mathbb{R}^8$.

- Mean center the data - substract a mean of intensity from each pixel. That is, two patches are now considered identical if their "brightness" differ by a constant. This puts all the patches into a 8-dimensional subspace of $\mathbb{R}^9$.
- Normalize with respect to the D-norm (normalize the contrast). That is, two patches are now considered identical if they only differ by "turning up contrast". This puts the data on a 7-dimensional ellipsoid inside $\mathbb{R}^8$.
- Do a coordinate change in order to make this ellipsoid into a 7-sphere S^7. Our data now is a point cloud $\mathcal{M}$ of approx. 4 millions points on S^7.

Analysis

How does $\mathcal{M}$ sit in S^7?

- Further heuristic: $\delta_k(x)$ is the distance of x from the k-th closest point of $\mathcal{M}$ to x.

How does $\mathcal{M}$ sit in S^7?

- Further heuristic: $\delta_k(x)$ is the distance of x from the k-th closest point of $\mathcal{M}$ to x.
- $\mathcal{M}[k, T] = \{x \in \mathcal{M} \mid$
  $\delta_k(x)$ is among the $T\%$ lowest values of δ_k on $\mathcal{M}\}$.

Analysis

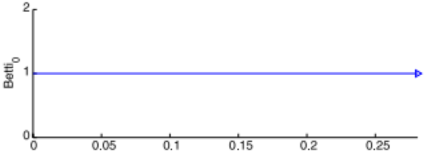
How does $\mathcal{M}$ sit in S^7?

- Further heuristic: $\delta_k(x)$ is the distance of x from the k-th closest point of $\mathcal{M}$ to x.
- $\mathcal{M}[k, T] = \{x \in \mathcal{M} \mid$ $\delta_k(x)$ is among the $T\%$ lowest values of δ_k on $\mathcal{M}\}$.
- For computational reasons, take samples of size approx. 10^4 from $\mathcal{M}$, and construct simplicial complexes using the witness method.

How does $\mathcal{M}$ sit in S^7?

- Further heuristic: $\delta_k(x)$ is the distance of x from the k-th closest point of $\mathcal{M}$ to x.
- $\mathcal{M}[k, T] = \{x \in \mathcal{M} \mid$ $\delta_k(x)$ is among the $T\%$ lowest values of δ_k on $\mathcal{M}\}$.
- For computational reasons, take samples of size approx. 10^4 from $\mathcal{M}$, and construct simplicial complexes using the witness method.
- Compute the persistent homology barcodes with coefficients in $\mathbb{Z}_2$ using the PLEX package.

Barcode in homological dimension 0 of $\mathcal{M}[15, 30]$:



Interpretation: $\mathcal{M}$ is connected (has one component).

Barcodes

Barcode in homological dimension 1 of $\mathcal{M}[300, 30]$ (very crude approximation):

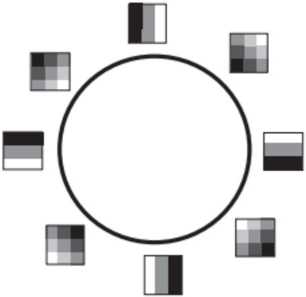
Interpretation: $\mathcal{M}$ in this detail has "1 hole" - simplest explanation is, that it looks like a circle.

Barcodes

Barcode in homological dimension 1 of $\mathcal{M}[300, 30]$ (very crude approximation):

Interpretation: $\mathcal{M}$ in this detail has “1 hole” - simplest explanation is, that it looks like a circle.

The picture below gives such an explanation.



Barcodes

Barcode in homological dimension 1 of $\mathcal{M}[15, 30]$:

Barcodes

Barcode in homological dimension 1 of $\mathcal{M}[15, 30]$:

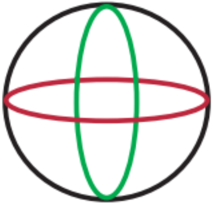
Interpretation: $\mathcal{M}$ actually has more 1-holes - the first Betti number seems to 5.

Barcodes

Barcode in homological dimension 1 of $\mathcal{M}[15, 30]$:

Interpretation: $\mathcal{M}$ actually has more 1-holes - the first Betti number seems to 5.

Carlsson et al. propose the following "3-circle" model:

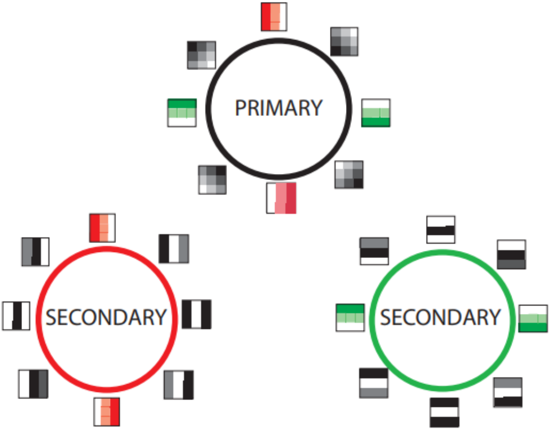


The green and the red circle both intersect the black one in two points. The red and the green circle are disjoint.

First Betti number = #arcs - #vertices + #components

=8-4+1=5

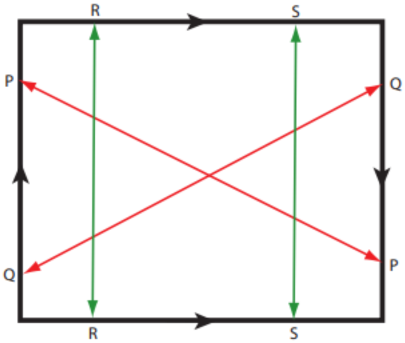
Again, a sort of explanation:



Primary circle corresponds to "linear patches". Two secondary circles correspond to "horizontally/vertically aligned patches". Take polynomial in 2 variables, and evaluate on set $\{-1,0,1\} \times \{-1,0,1\}$. Linear, quadratic patches...

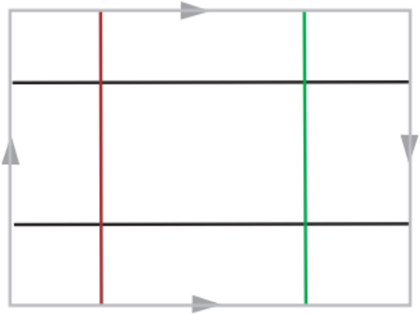
Klein bottle

Klein bottle is a 2-surface defined as a topological quotient of a square:



Klein bottles does not embed into $\mathbb{R}^3$, but it does embed into $\mathbb{R}^4$.

The 3-circle model embeds into the Klein bottle:



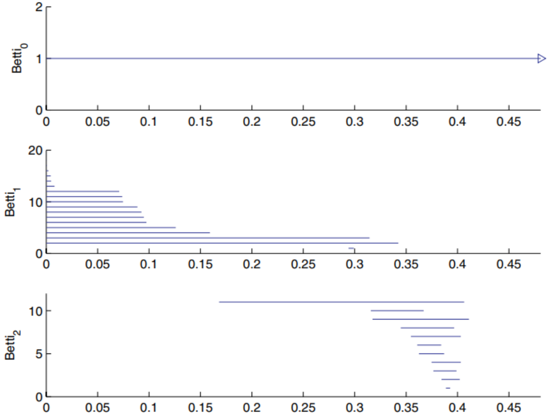
- What now? Meddle with the parameters until we find the Klein bottle.

- What now? Meddle with the parameters until we find the Klein bottle.
- The Betti numbers with coefficients in $\mathbb{Z}_2$ of the Klein bottle are $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1,\ \beta_i = 0$ for $i > 2$.

- What now? Meddle with the parameters until we find the Klein bottle.
- The Betti numbers with coefficients in $\mathbb{Z}_2$ of the Klein bottle are $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$, $\beta_i = 0$ for $i > 2$.
- While changing parameter k, we unfortunately always get zero second Betti number.

- What now? Meddle with the parameters until we find the Klein bottle.
- The Betti numbers with coefficients in $\mathbb{Z}_2$ of the Klein bottle are $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1,\ \beta_i = 0$ for $i > 2$.
- While changing parameter k, we unfortunately always get zero second Betti number.
- More heuristic to the rescue... Authors argue that their method has preference for linear and vertically/horizontally aligned patches. They add a certain set of points $\mathcal{Q}$ from $\mathcal{M}$ not in $\mathcal{M}[100, 30]$ of pure quadratic character.

Barcodes for $\mathcal{M}[100, 30] \cup \mathcal{Q}$:



Interpretation: Point cloud $\mathcal{M}[100, 30] \cup \mathcal{Q}$ "looks like" something with Betti numbers $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$.

More handwaving: There are two 2-surfaces with Betti numbers mod 2 being $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$ - Klein bottle and the torus. Their homology groups differ mod 3 though, and the authors claim to have done the corresponding persistent homology computation, and that it is in the favor of the Klein bottle...


BE PERSISTENT!
wikiHow to Be Persistent