Persistent cohomology and circular coordinates

Mikael Vejdemo-Johansson Vin de Silva Dmitriy Morozov

June 27, 2010

Point clouds

Features of modern data analysis

- High dimensionality
- Large datasets
- Increasing interest in subtle relationships

Dependencies and relationships in data expressible as geometric concepts.

Often, data is expected to be a (noisy) sample of points on some manifold in some (high-dimensional) ambient space.

Quantity vs. Quality

Classically

- Data analysis and statistics concerned with simple descriptors
- Mean, median, variance, percentiles, ...
- Focus on quantitative measurements
- Only works if metrics are easily justified (physics: easy, genetics: harder)

Topologically

- Metrics no longer trusted: instead interested in connectivity and closeness.
- Less interested in quantitative (how large variance), and more interested in qualitative (how many components) properties.

Approach at Stanford

Fundamental object to study: finite metric space $(\mathbb{X}, \mu)$, preferably with embedding in ambient space $\mathbb{R}^d$.

Geometric properties captured by topological data:

Čech complex

$C(\mathbb{X}, \epsilon)$ contains simplex $v_0, \ldots, v_k$ if there is a point v in ambient Euclidean space such that all $\mu(v_i, v) < \epsilon$.

Computationally easier:

Vietoris-Rips complex

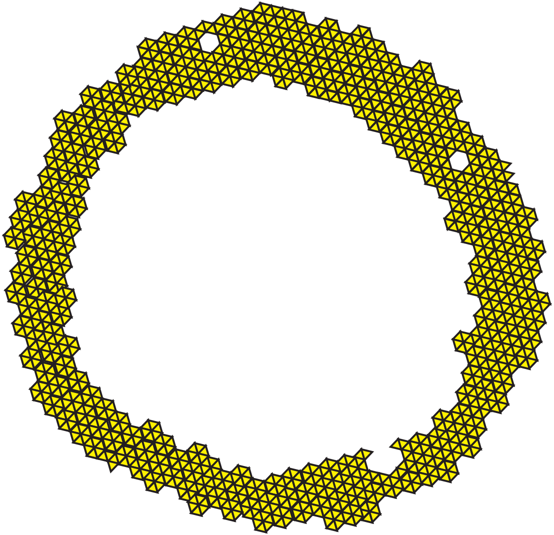
$C(\mathbb{X}, \epsilon)$ contains simplex $v_0, \ldots, v_k$ if all $\mu(v_i, v_j) < \epsilon$.

As ϵ varies, the Čech and Vietoris-Rips complexes capture topological features of a point cloud at different scales.

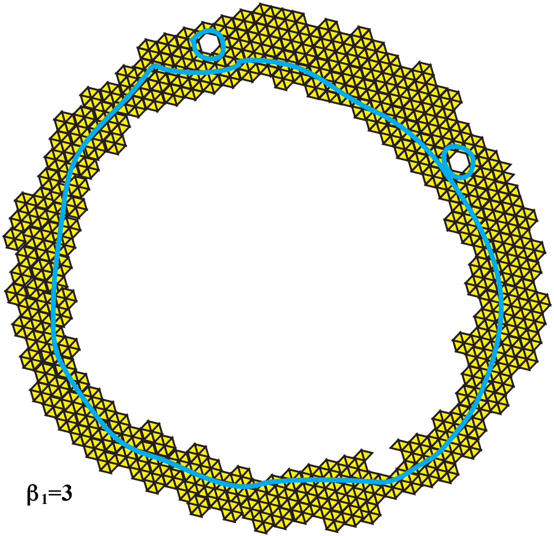
Scale dependent features

- Choice of ϵ determines scale
- Different scales highlight different topological features.
- How do we choose ϵ?
- How do we tell features from noise?

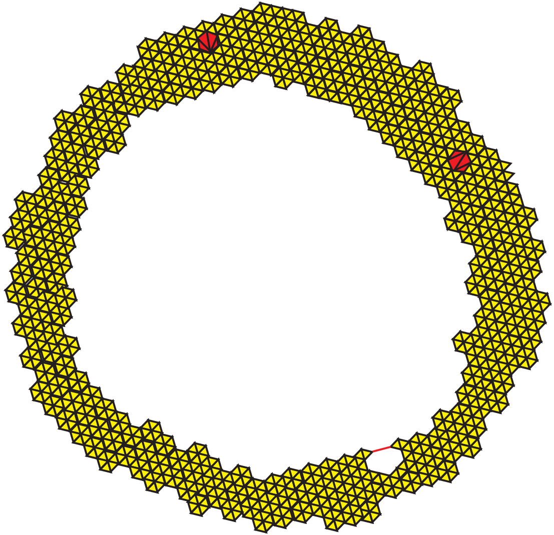
Scale dependent features



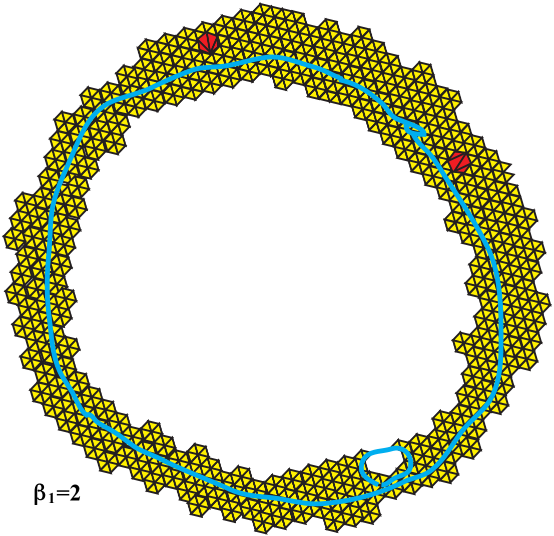
Scale dependent features



Scale dependent features



Scale dependent features



Persistence

Vietoris-Rips and Čech complexes yield filtered simplicial complexes: $C(\mathbb{X}, \epsilon_0) \subseteq C(\mathbb{X}, \epsilon_1)$ if $\epsilon_0 < \epsilon_1$.

We call *persistent* from ϵ to ϵ' those representative cycles in $H_k(C(\mathbb{X}, \epsilon_1))$ that lie in the image of the homomorphism $H_k(C(\mathbb{X}, \epsilon_0)) \to H_k(C(\mathbb{X}, \epsilon_1))$ induced by the inclusion.

Algebraic framework

Chain complex from (finite) filtered simplicial complex: graded $k[t]$-module. t acts by inclusion from one filtration step to the next.

Homology is a subquotient of $k[t]$-modules. t-action carries by functoriality.

If coefficients lie in a field, structure theorems for PIDs yield:

$$H = \bigoplus \Sigma^{d_i} k[t] \oplus \bigoplus \Sigma^{d_i} k[t]/t^{d_j}$$

We capture the decomposition as a barcode: an assembly of intervals $[d_i, \infty)$ and $[d_i, d_j)$.

Length of interval corresponds to importance of the feature it describes.

Persistence applications

To date, a number of applications have been found for persistent homology:

- Diabetes type I and II easily recognizable with topological clustering. [Carlsson et.al.]
- Cell phone coverage quality can be done by local homology computations. [de Silva, Ghrist]
- 3×3 pixel patches from natural images distribute on a Klein bottle. [Carlsson, Ishkanov, de Silva, Zomorodian]
- Textures can be characterized by the density distribution on this Klein bottle. [Carlsson, Perea]

Coordinate functions give quantitative tools

Classically

Powerful tools for data analysis given by Principal Component Analysis and Singular Value Decomposition.

- Linear change of basis for space of data
- New basis exhibits interesting features in small subspaces
- Basis vector v corresponds to coordinate function $c_v : X \to \mathbb{R}$ given by $c_v(d) = \langle v, d \rangle$.

Moral: coordinate functions are helpful for data analysis

New directions

Find coordinate functions.

- Looser requirements: topological maps, not necessarily linear
- Wider coordinate spaces: drop requirement of real-valued coordinates.

What is the next easiest coordinate space?

Automating coordinate recovery

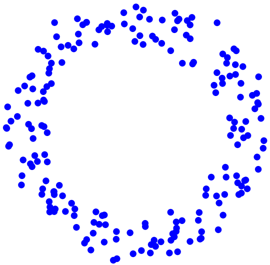
In order to automate the process used for the Klein bottle, we are studying topological techniques to recover coordinate functions.

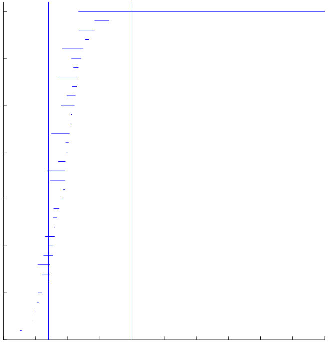
In (de Silva – Morozov – V-J, 2009), we propose using the natural isomorphism of functors

$$H^1(-, \mathbb{Z}) \cong [X, S^1]$$

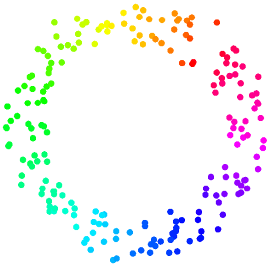
to produce circular coordinates. This requires a persistent cohomology framework to produce good coclasses, and a smoothing step to produce useful coordinate functions.

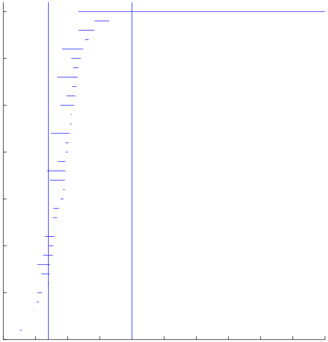
Persistent cohomology and circular coordinates



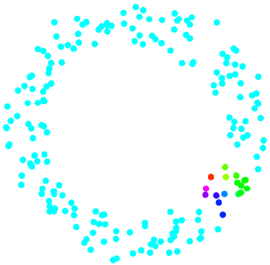


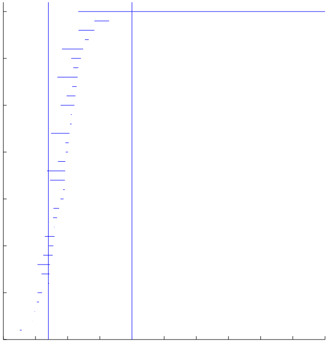
Persistent cohomology and circular coordinates



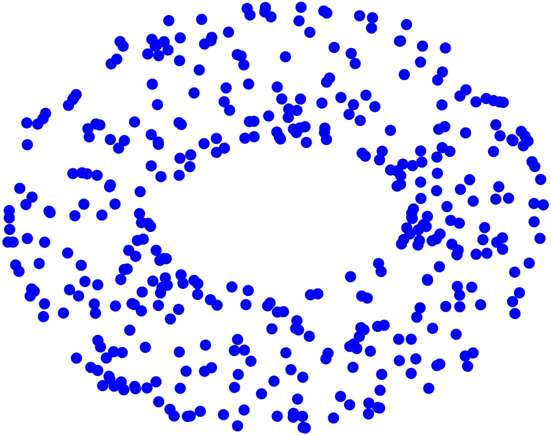


Persistent cohomology and circular coordinates

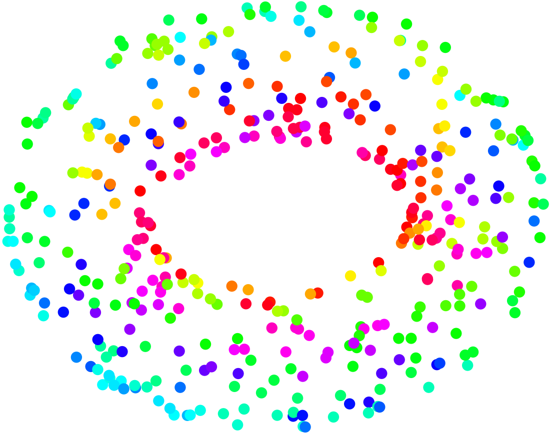




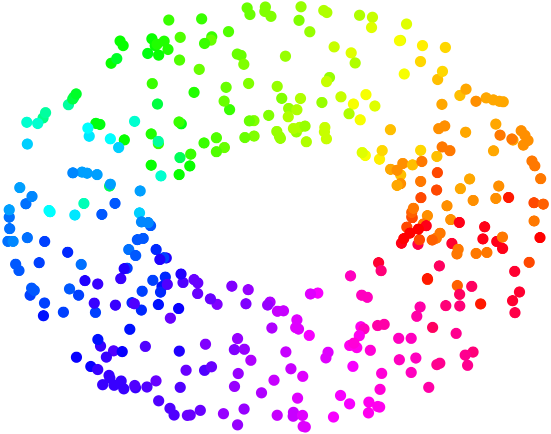
Persistent cohomology and circular coordinates



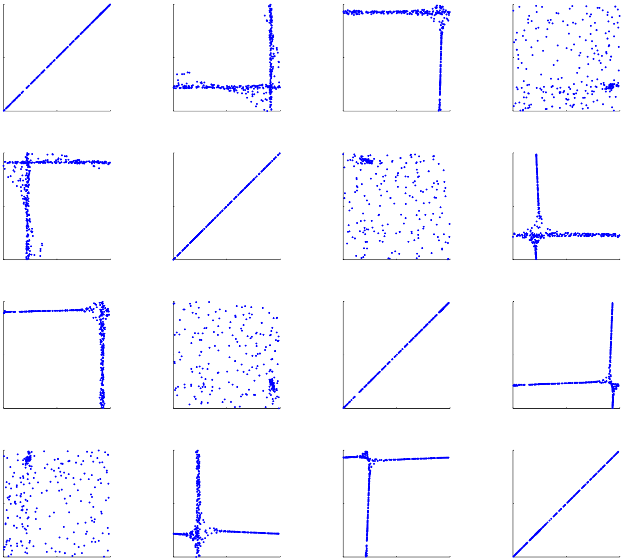
Persistent cohomology and circular coordinates



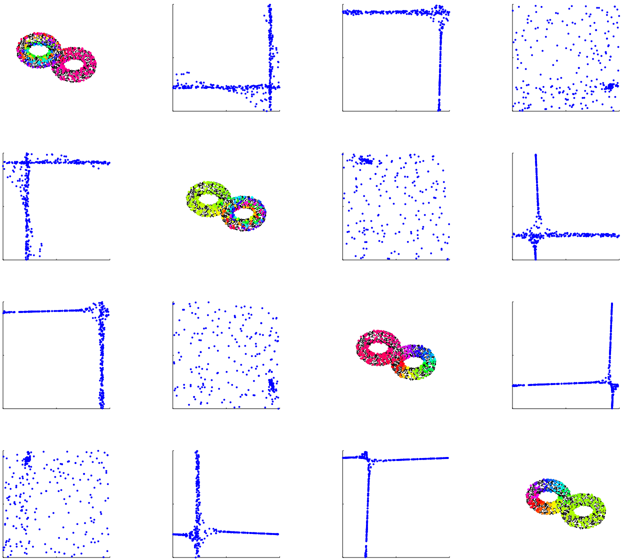
Persistent cohomology and circular coordinates



Persistent cohomology and circular coordinates



Persistent cohomology and circular coordinates



Future directions

Mixed coordinate systems

- We can work over open cubes (classically) and over tori ($S^1 \times S^1 \times \cdots \times S^1$).
- Would be interesting to combine different coordinates
- Working on using $H_0(\hom(C_*X, C_*Y), d_Y f \pm f d_X)$ to produce coordinates for X in Y for arbitrary simplicial complexes.

Periodicity analysis

- Single non-trivial 1-cycle might indicate recurrence
- Coordinatization yields intrinsic progression for recurrent systems
- Yields approach to periodic systems fundamentally different from recurrence diagrams and from fourier analysis