Persistent cohomology and circle-valued coordinates: analyzing periodicity with sparse sampling

Mikael Vejdemo-Johansson

September 1, 2009

Outline

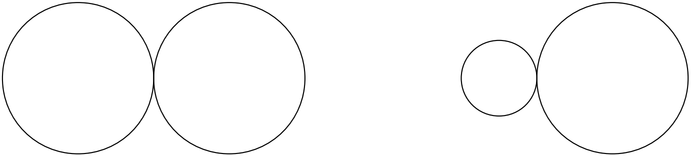
A primer on algebraic topology

Persistent algebraic topology

Cohomology and circular coordinates

Geometry without metrics

- Topology is about geometry, but without reliable metric.
- Fundamental notion of topology: connectivity.



- Different geometries — sizes different.
- Same topology — connectivity of points on the figures similar.
- Topological analysis easy even if metrics are non-existent or non-natural.

This all allows us to find invariants measuring qualitative properties (i.e. there are two circles) without getting distracted by quantitative properties (i.e. their sizes).

Homotopy

- Algebraizing topology by looking at *groups* of continuous maps from spheres to the topological space.

+ Precise tool.

– Difficult to handle. Even the homotopy of the spheres are not known in full detail.

- An element is an equivalence class of maps from a given sphere to the space, under “small changes”. See path-invariance of integrals.

Suppose X is a topological space.

$\pi_0(X)$ The zero'th homotopy “group”. Elements correspond to connected components of X.

$\pi_1(X)$ Elements are closed paths, wrapped around “holes”.

(Co-)Homology

- Algebraizing topology by deriving questions about linear maps between vector spaces from the topological questions.
- Fundamental entity: $\ker D/\operatorname{im} D$, for a linear operator D such that $D^2 = 0$.
- Homology arises naturally from cell decompositions of spaces. Homology elements correspond to locations of holes or bubbles.
- Cohomology arises naturally from Stokes theorem and differential forms. Cohomology elements correspond to failure modes for Stokes theorem.
- Dual theories — tight correspondences. For good spaces.
- $\beta_i = \dim_k H_i(X; k)$ the Betti numbers. Measure rank of homology groups.
- β_0 — counts connected components.
- β_1 — counts holes.

Outline

Basic scenario

Much of the research and methods in persistent algebraic topology run along the same lines of inquiry:

1. Data present as a "point cloud": a finite, possibly large subset $X \subset \mathbb{R}^d$ (or just a finite metric space, possibly).
2. This point cloud is assumed to be a *good* sampling from some manifold.
3. By creating topological entites from this point cloud, we strive to determine topological invariants for the underlying manifold in a way that is resistant to noise, and gives us confidence margins.

Generally, we can produce the topological invariants dependent on some parameter; and the *persistence* part gives a framework to sweep through all relevant parameter values and get an overall picture for further analysis.

Encoding metrics without metrics

For any given *threshold* value ε, we can create a graph from our data points, inserting edges whenever points are less than ε apart. This gives us a including sequence of graphs encoding the metric data inherent in the original points, since with larger ε we gain new edges, but never lose old edges.
Several ways to build a topology out of a distance graph:

- Čech complex: connect *k* points whenever they all are within ε.
- Vietoris–Rips complex: connect *k* points whenever they are pairwise within ε.
- α-shapes: form substructures of a Delaunay triangulation induced by the graph.
- Witness complexes: connect *landmark points* whenever they are closely connected by other data points.

Persistence

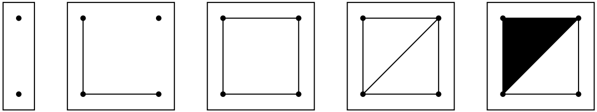
Homology and homotopy both are *functors*, which means that a map of spaces $X \to Y$ induces a map of homology, cohomology or homotopy groups.
Thresholding gives us a chain of inclusion maps

$$\ldots \hookrightarrow X_{\varepsilon_1} \hookrightarrow X_{\varepsilon_2} \hookrightarrow X_{\varepsilon_3} \hookrightarrow X_{\varepsilon_4} \hookrightarrow \ldots$$

which gives us a chain of, say, homology groups, or Betti numbers, or homotopy groups, and so on.
Finding good structure descriptions for these gives us powerful analysis tools.

Betti numbers and barcodes

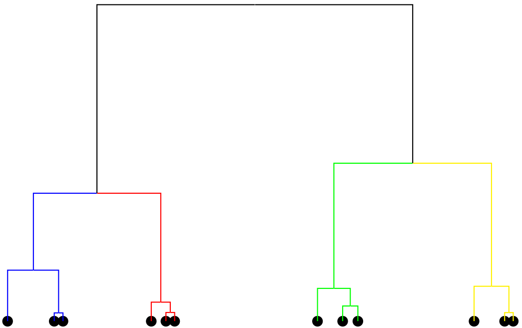
Analysing these chains with Betti numbers gives us a first glimpse of the methods used:

t	0	1	2	3	4
X					
β_1	0	0	1	2	1
β_0	2	2	1	1	1

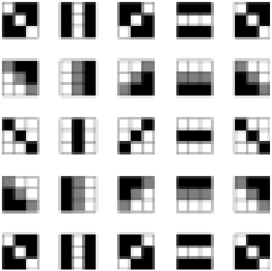
Corresponding barcodes drawn on board.

Application: Clustering and Dendrograms

It turns out that doing π_0 with persistence gives us something statisticians are already using quite a lot: Dendrograms!



Application: Image spaces



Consider natural images. These are composed out of pixels. Picking out 3x3 pixel configurations, we get building blocks of natural images — as vectors in $\mathbb{R}^9$.
[Mumford et al.: 1999] compiled a large data set with high contrast 3x3-patches from natural images, normalized to land on a 7-sphere in $\mathbb{R}^8$.
Persistent algebraic topology gave hints that were then synthesized, in [Carlsson, Ishkanov, de Silva, Zomorodian: 2008] to a description of the space of such image patches as a Klein bottle, leading to a new algorithm for image compression.

Outline

Useful topology

Taking a leaf from algebraic geometry, we view a *coordinate* as a function from the space to a *coordinate space*.

Cohomology as functions

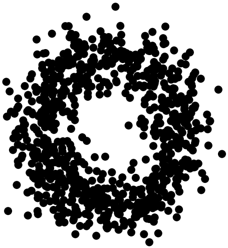
A fundamental fact from algebraic topology is:

$$H^1(X;\mathbb{Z}) = [X, S^1] = \{f : X \to S^1 \mid \text{up to homotopy}\}$$

Thus, computing cohomology is *the same* as computing a circle-valued coordinate function for the space.

Circle parametrizations

We have code up and running to do this analysis, and to construct circle-valued functions on data sets.



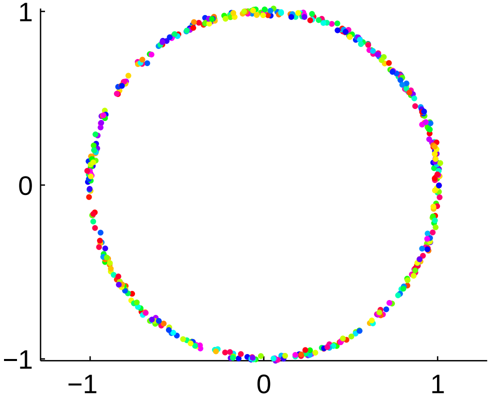
Periodicity as circular paths

For nice enough periodic dynamical systems, the path in phase space a single orbit traverses forms, topologically, a circle. We have been able to use the circular parametrization techniques here to recover period lengths without introducing additional noise the way difference-based methods would.

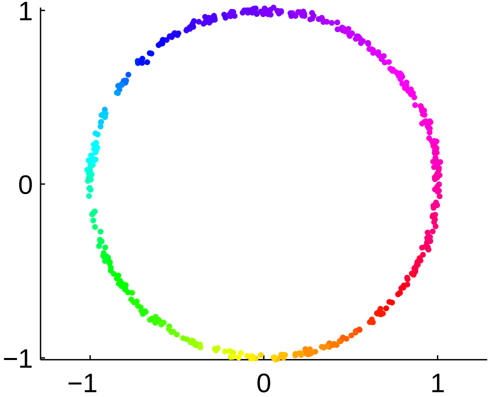
Taking a large enough sample — either densely sampled, or sparsely but over long time — from a time series, and ignoring, at first, the time coordinate, we acquire a circular coordinatization. This coordinatization then is *unrolled*, by looking for places where the coordinate has wrapped around the circle, and increasing an added offset.

This way, we get a time vs. coordinate correspondence with inclination the period length.

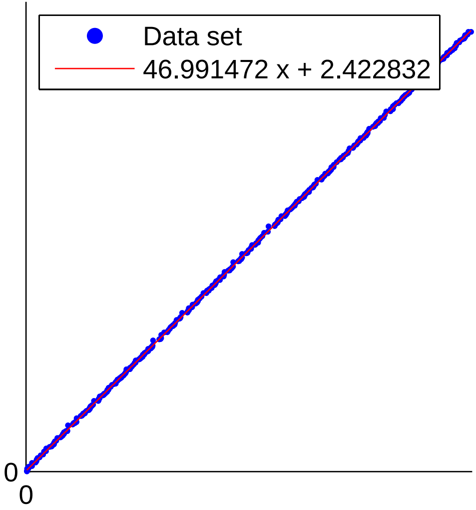
A first, ideal example



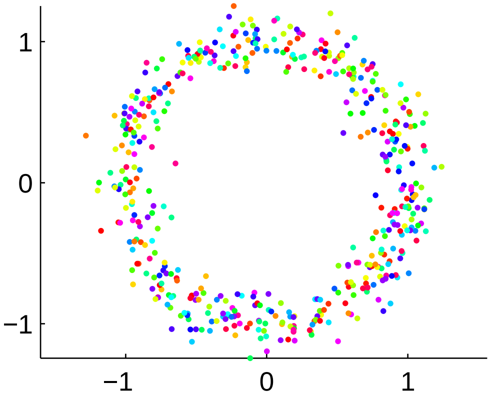
A first, ideal example



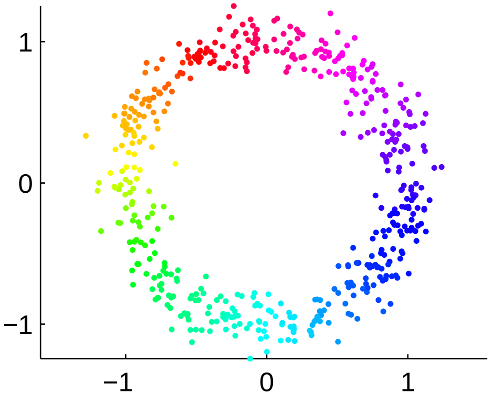
A first, ideal example



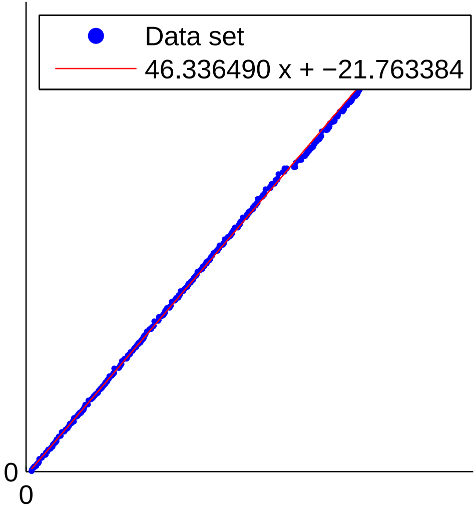
Adding more spatial noise



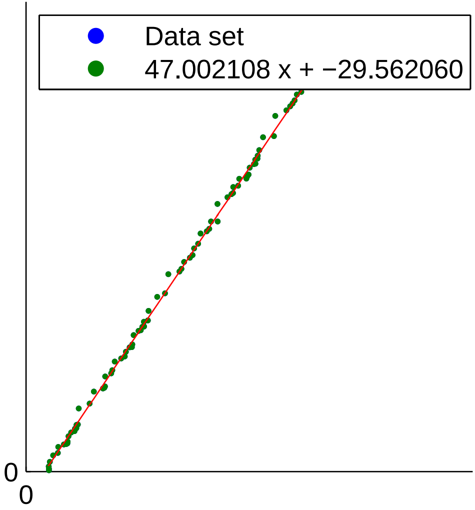
Adding more spatial noise



Adding more spatial noise



Adding more spatial noise



Future directions?

And here we come to why I'm here today:

Where can we go from here?

What could we do with dynamical systems, if we had a topologically stable way of finding circular coordinate functions?