Persistent Cohomology and Circle-valued coordinates

Mikael Vejdemo-Johansson Vin de Silva

April 1, 2009

Översikt

1. Motivation: Inherenta koordinater
2. Teori: Persistent kohomologi och cirkelvärda funktioner
3. Praktik: Att hitta och tolka koordinatiseringar

Introduktion och motivation

1. Drivkraft: hitta inneboende koordinater till datapunkter
2. Klassiska koordinatmängder för begränsade: cirkeln för kognitivt överlastad – tar upp 2 koordinater istället för 1.
3. Istället: utvidga koordinatfunktionerna från linjära till linjära+cirkelvärda
4. Effektiva algoritmer: variera persistensalgoritmen.

Perspektiv och plats i den stora helheten

Det är alltid viktigt, och svårt, att hitta en plats för varje arbete i sin kontext, i sin helhet. Än mer så när man rör sig utanför de hemvana vidderna, i ett fält man agerar som nykomling i.

Perspektiv och plats i den stora helheten

Det är alltid viktigt, och svårt, att hitta en plats för varje arbete i sin kontext, i sin helhet. Än mer så när man rör sig utanför de hemvana vidderna, i ett fält man agerar som nykomling i.Därav kommer detta arbete att förlita sig på att dataloger tar hand om de viktigare frågorna, såsom till exempel

$$P = NP \qquad P \neq NP$$

vilka säkerligen har djupa efterverkningar även på detta arbete. Med ett tillräckligt kraftfullt orakel blir våra algoritmer snabba nog att man kan få ner komplexitetsklassen bortom vad den bevisbart befinner sig vid.

Perspektiv och plats i den stora helheten

Det är alltid viktigt, och svårt, att hitta en plats för varje arbete i sin kontext, i sin helhet. Än mer så när man rör sig utanför de hemvana vidderna, i ett fält man agerar som nykomling i.Därav kommer detta arbete att förlita sig på att dataloger tar hand om de viktigare frågorna, såsom till exempel

$$P = NP \qquad P \neq NP$$

vilka säkerligen har djupa efterverkningar även på detta arbete. Med ett tillräckligt kraftfullt orakel blir våra algoritmer snabba nog att man kan få ner komplexitetsklassen bortom vad den bevisbart befinner sig vid.Ytterligare kulturhänsyn måste tas, med temporal sensitivitet, till saker som datum. Då de här föredraget hålls på Första April, kan jag ju knappast undvika den gamla goda traditionen av aprilskämt, eller som de heter på engelska:

April Fool's

And now, for something completely different.

Outline

Motivation: Intrinsic coordinates

Theory: Persistent cohomology and circle-valued maps

Practice: Finding and interpreting coordinatizations

Finding coordinates

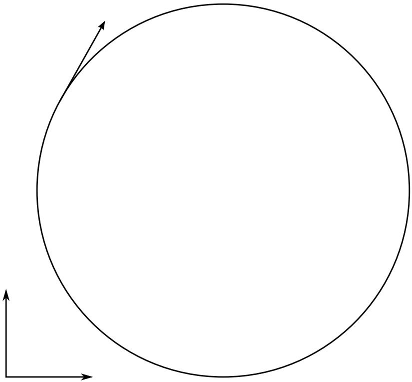
- Overall goal is to understand and analyze datasets.
- Data comes with coordinates.
  Different coordinate choice might concentrate intrinsic information.
- Want: find few and very relevant intrinsic coordinates.

Problematic cases

Some shapes take up too many coordinates

Problematic cases

Some shapes take up too many coordinates



Locally 1-dimensional. Globally 2 coordinates needed to describe all points. The shape doesn't fit in $\mathbb{R}$.

Similar problems arise with sphere and torus.

Fixes

How can we fix this?

Circle-valued coordinates

- Use $S^1 = [0, 1]/(0 \sim 1)$ as coordinate space
- Fixes the circle
- Fixes the torus
- Occurs naturally:
  - Phase coordinates for waves
  - Angle coordinates for directions

Outline

Motivation: Intrinsic coordinates

Theory: Persistent cohomology and circle-valued maps

Practice: Finding and interpreting coordinatizations

Persistent cohomology

Problem remains: how do we find circle-valued coordinates?

Persistent cohomology

Problem remains: how do we find circle-valued coordinates?

Persistent Cohomology

- Degree one cohomology equivalent to circle-valued maps
- Persistence picks out relevant features
- Once a feature-rich parameter has been found, we work in ordinary (non-persistent) cohomology theories

The algorithm we use is a variation on the Persistence algorithm. We introduce simplices one after the other, and reduce a cumulative coboundary matrix. The same argument as for the Persistence algorithm gives us a complexity estimate of $O(n^3)$ in the number of simplices, but from experience we expect much better performance.

From cohomology to circle-valued coordinatizations

Use canonical isomorphism

$$H^1(X) \cong [X, S^1]$$

Issues

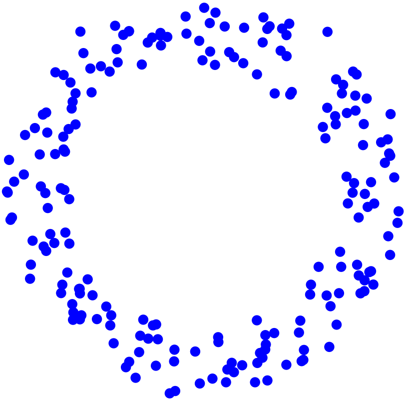
- Easy to compute: Modular cohomology, coefficients in $\mathbb{Z}/p$ for small primes p.
  Need for the isomorphism: Real-valued cohomology.
  Smoothness: Integral cohomology gives constant values on all vertices, and wraps edges in the complex around the target circle.
- Numerical stability of cohomology computation and of the smoothing operations.

Smoothing

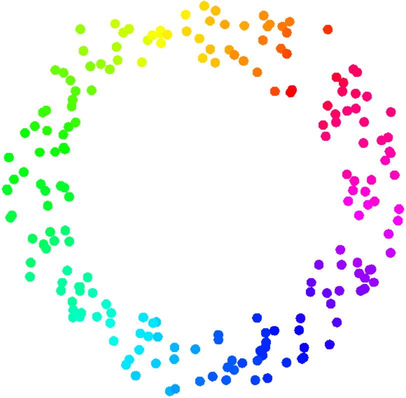
- Integral 1-cocycle: integer weighted edge graph.
- Coordinates found by edge traversal, increasing by edge weights.
- This algorithm guaranteed by cocycle condition to give the same values, mod 1.0, to each vertex.
- Application straight on integral cocycle yields value 0 at each vertex.
- Given ζ integral cocycle, we wish to find cohomologous cocycle z such that the edges are small.
- Hence, we wish to find x such that $\zeta + \partial x$ has minimal L_2-norm.
- This is a well-known optimization problem. We use the LSQR algorithm.

Outline

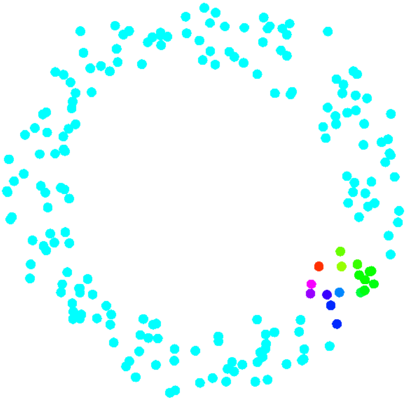
Parametrized circles



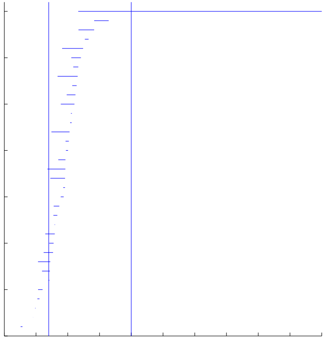
Parametrized circles

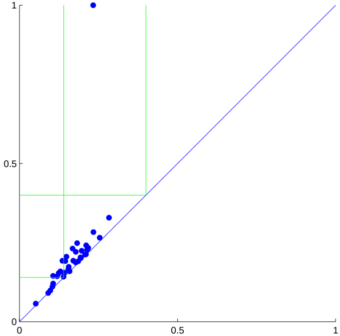


Parametrized circles

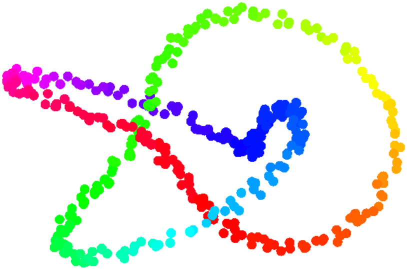


Parametrized circles

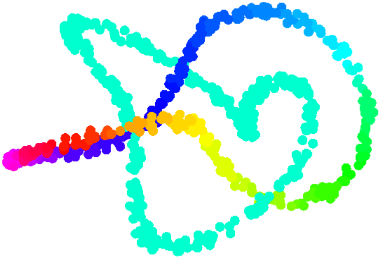




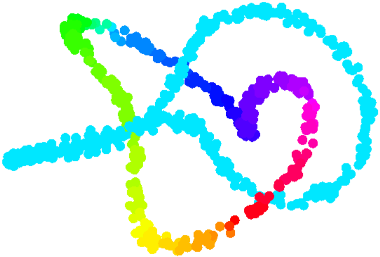
Knots and links



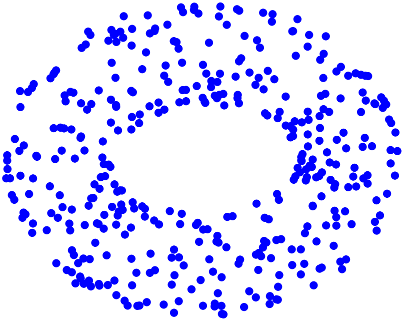
Knots and links



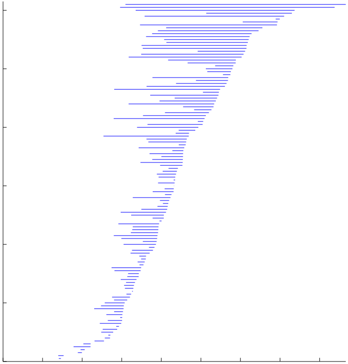
Knots and links



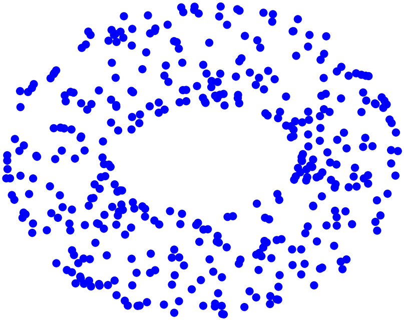
Torus



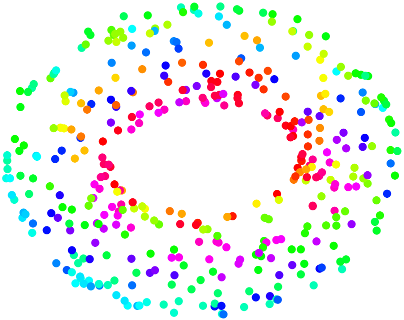
Torus



Torus

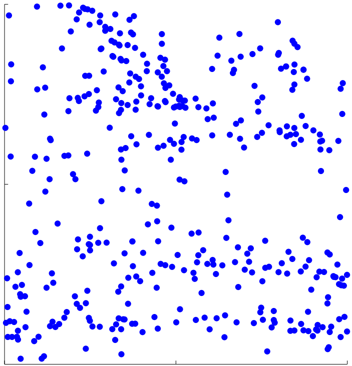


Torus



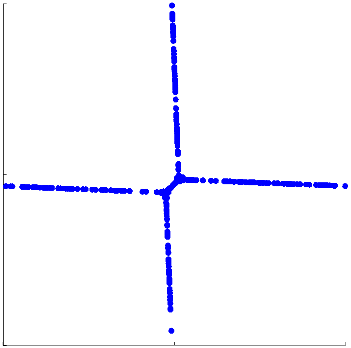
Torus

Torus



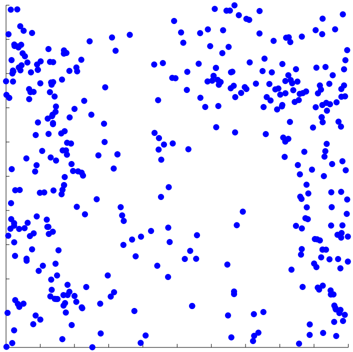
Correlation plot for this torus parametrization

Torus



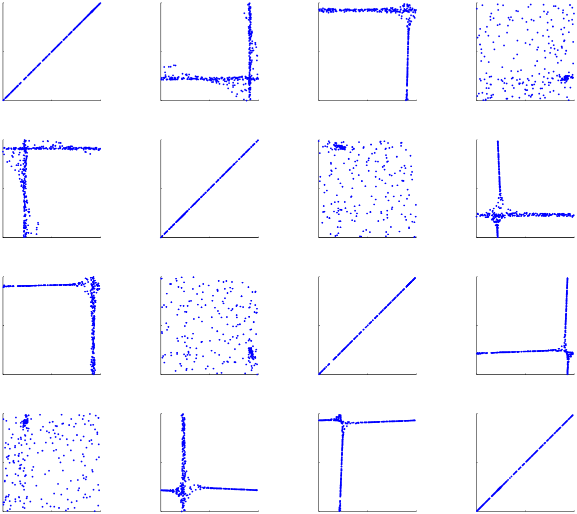
Correlation plot for a wedge of two circles

Torus

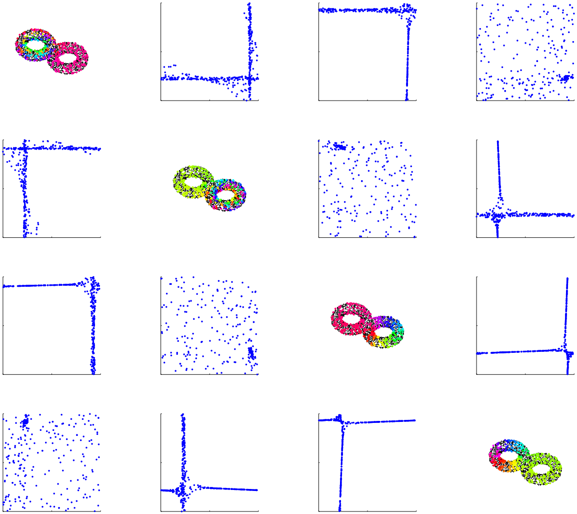


Correlation plot for an elliptic curve in $\mathbb{C}P^2$

Pop quiz



Pop quiz – the Double Torus



Mumford dataset

Lee-Mumford-Pedersen, *The nonlinear statistics of high-contrast patches in natural images*, International Journal of Computer Vision 54(1/2/3), 83-103, 2003

$4.2 \cdot 10^6$ pixel patches from 4167 calibrated 1020×1532 images.

Each 3×3 pixel patch obviously a vector in $\mathbb{R}^9$. Normalized to constant intensity and to unit euclidean norm.

Transformed by a basis choice that highlights geometric features of the dataset itself.

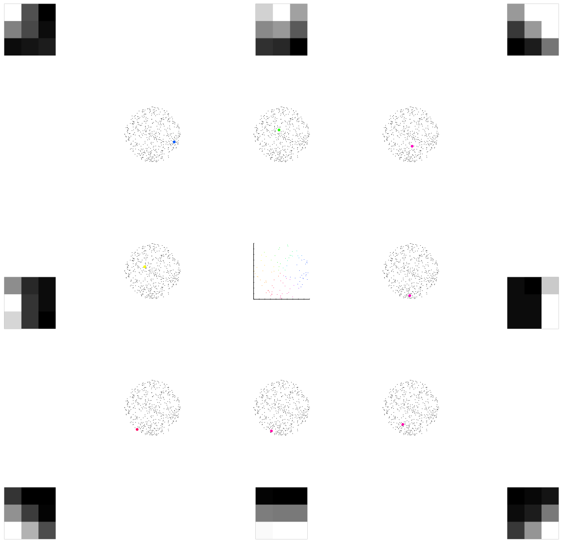
Result lies on the unit 7-sphere in $\mathbb{R}^8$.

Mumford dataset

We use the smoothing procedure developed by Jennifer Kloke. Once smoothed to a circle, we parametrize with persistent cohomology, and can pull the parametrization back to the original data points.

Movie Time!

Mumford dataset



Acknowledgements

Thanks are due for this to:

- Vin de Silva, Dmitriy Morozov – my collaborators
- Jennifer Novak Kloke – smoothed Mumford data
- Gunnar Carlsson
- BIRS and the organizers
- ONR, DARPA-TDA, Pomona College and Stanford University – funding