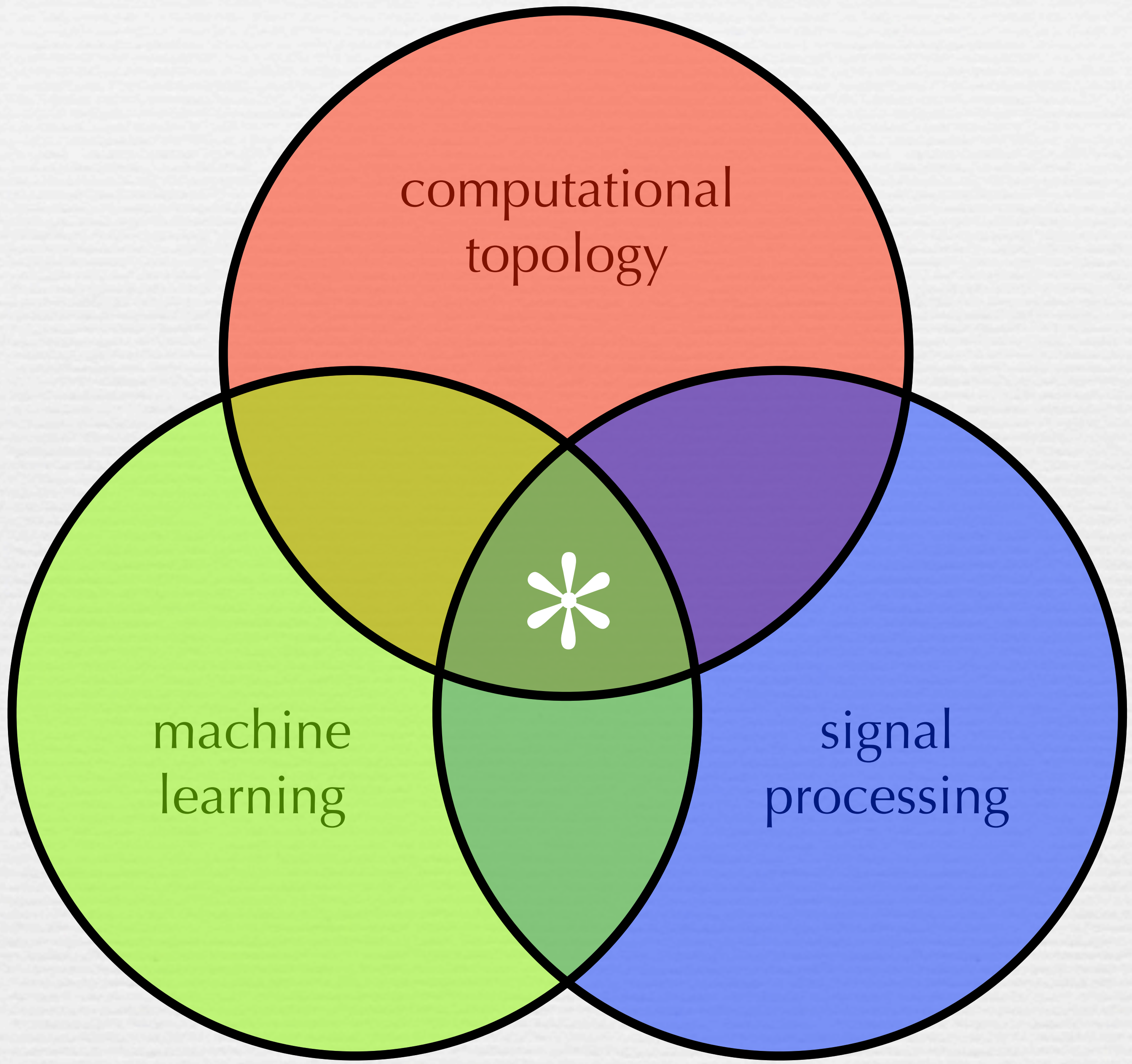
Topological Dimensionality Reduction

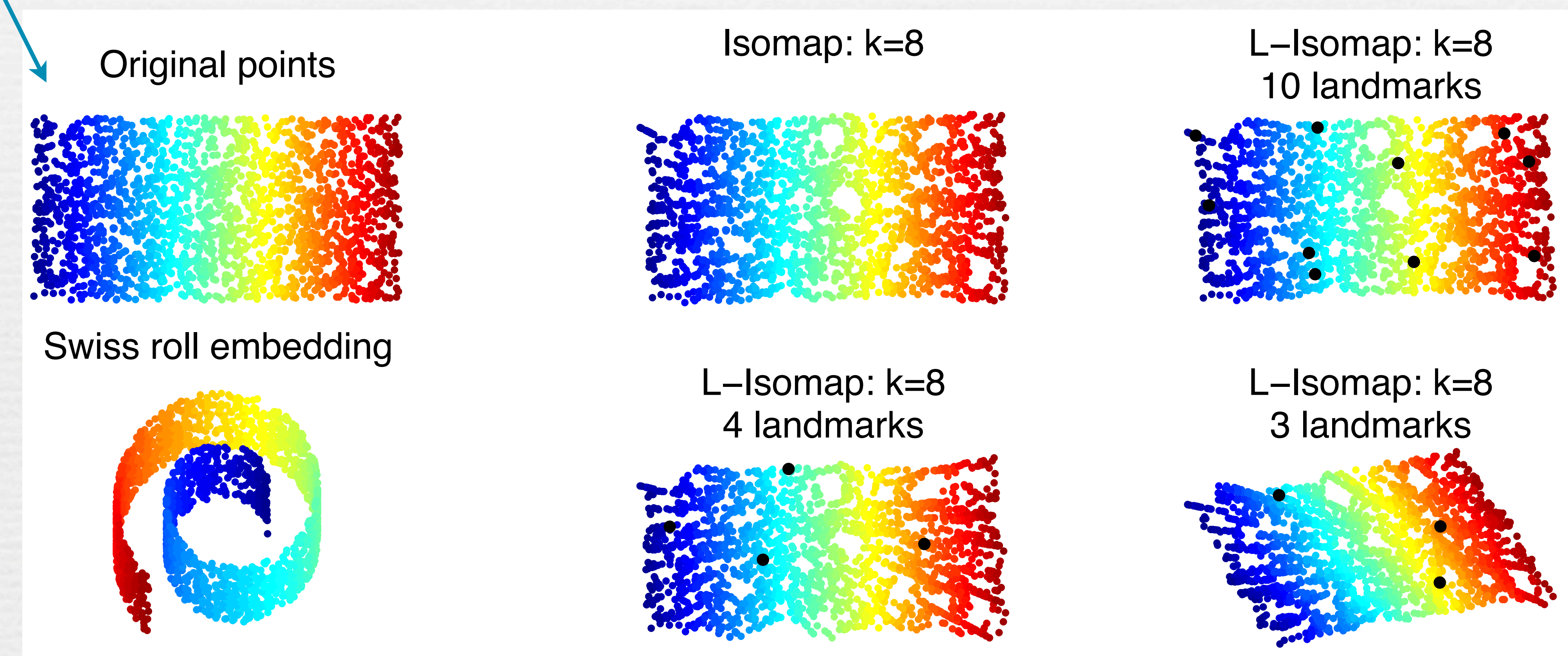
Vin de Silva, Primoz Skraba, Mikael Vejdemo-Johansson



Nonlinear dimensionality reduction

Nonlinear dimensionality reduction

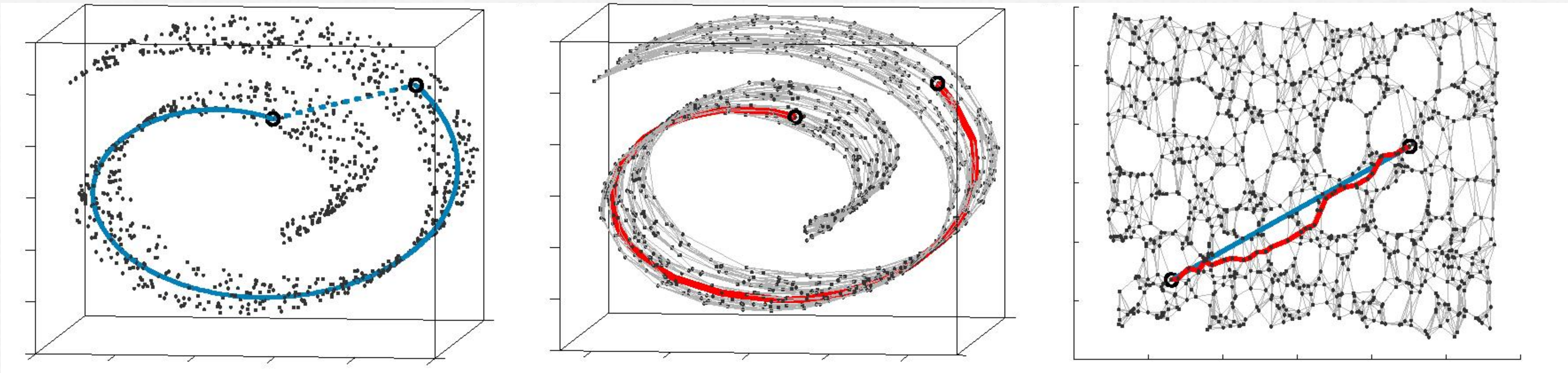
unknown: linear parameter space



input: nonlinear observed data

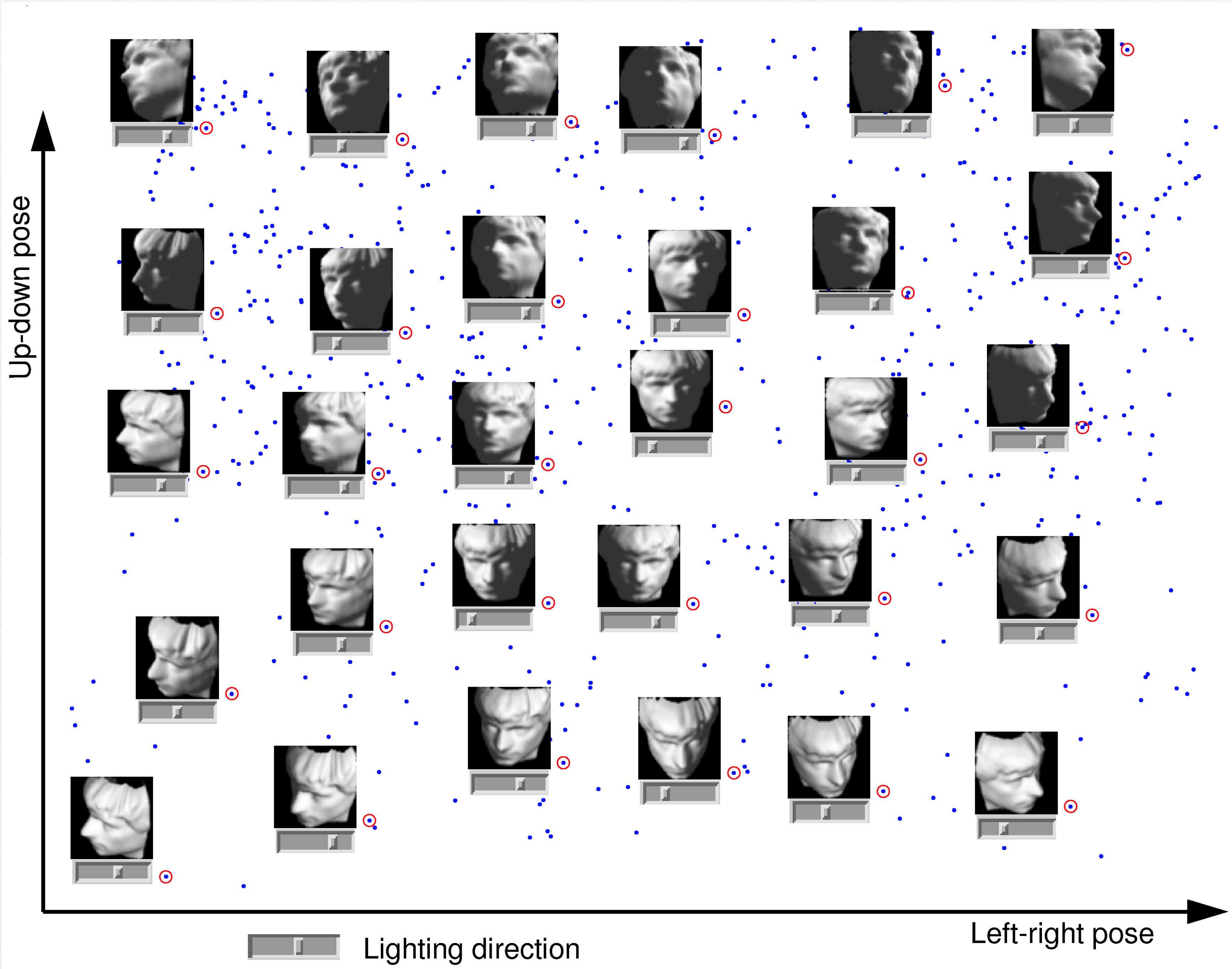
output: low-dimensional coordinate embedding

How Isomap works



- True distance measured as geodesics along the surface (left)
- Surface geodesics approximated by graph geodesics (middle)
- Input graph geodesic distances into classical MDS (multidimensional scaling) for coordinate embedding (right)

Example: face images



NLDR techniques

- Since December 2000:
  - Isomap (Tenenbaum, dS, Langford) — geodesics
  - LLE (Roweis, Saul) — local affine structure
  - Laplacian Eigenmaps (Belkin, Niyogi) — diffusion geometry
  - Hessian Eigenmaps (Donoho, Grimes) — 2nd fundamental form
  - ...

Laplacian Eigenmaps (Belkin & Niyogi)

- Represent data by graph, then:
  - cochain spaces

$$C^0 = \text{vector space spanned by vertices} \cong \{f : V \to \mathbb{R}\}$$ scalar fields

$$C^1 = \text{vector space spanned by edges} \cong \{\alpha : E \to \mathbb{R}\}$$ vector fields

  - coboundary map

$$\delta : C^0 \to C^1; \quad \delta f([ab]) = f(b) - f(a)$$ discrete gradient

(signed incidence matrix between edges and vertices)

  - discrete Laplacian

$$\Delta_0 = \delta^* \delta : C^0 \to C^0$$

(diagonal entries = -degree; off-diagonal entries 0 or -1)

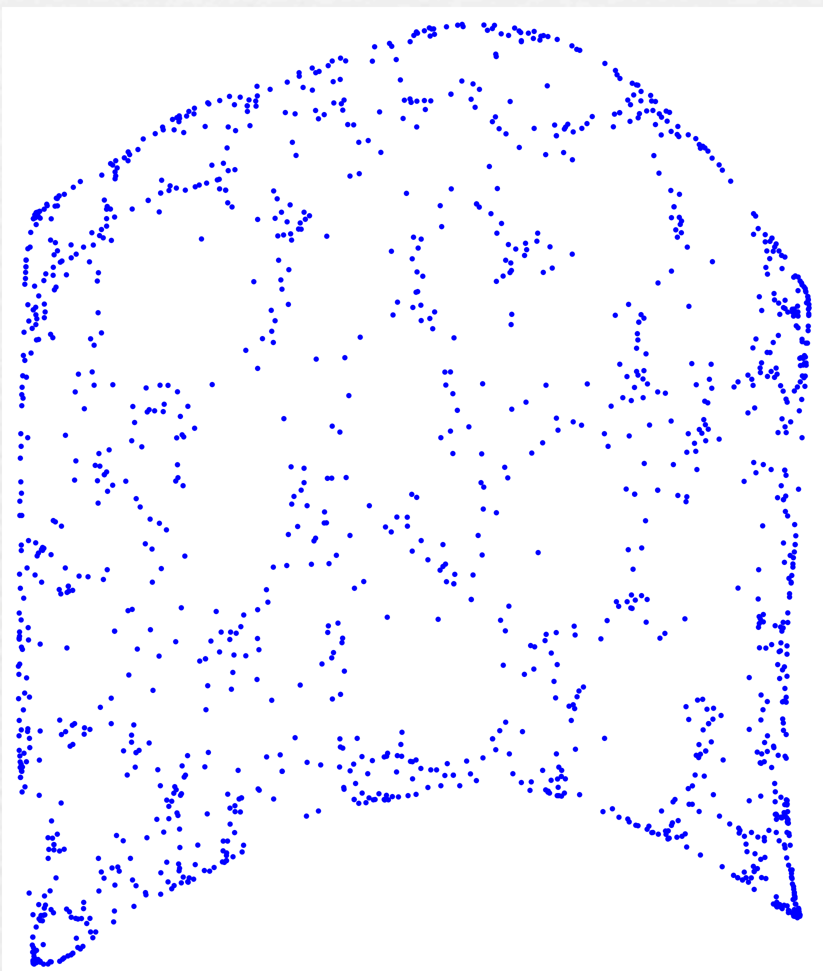
  - eigenvalues

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots$$

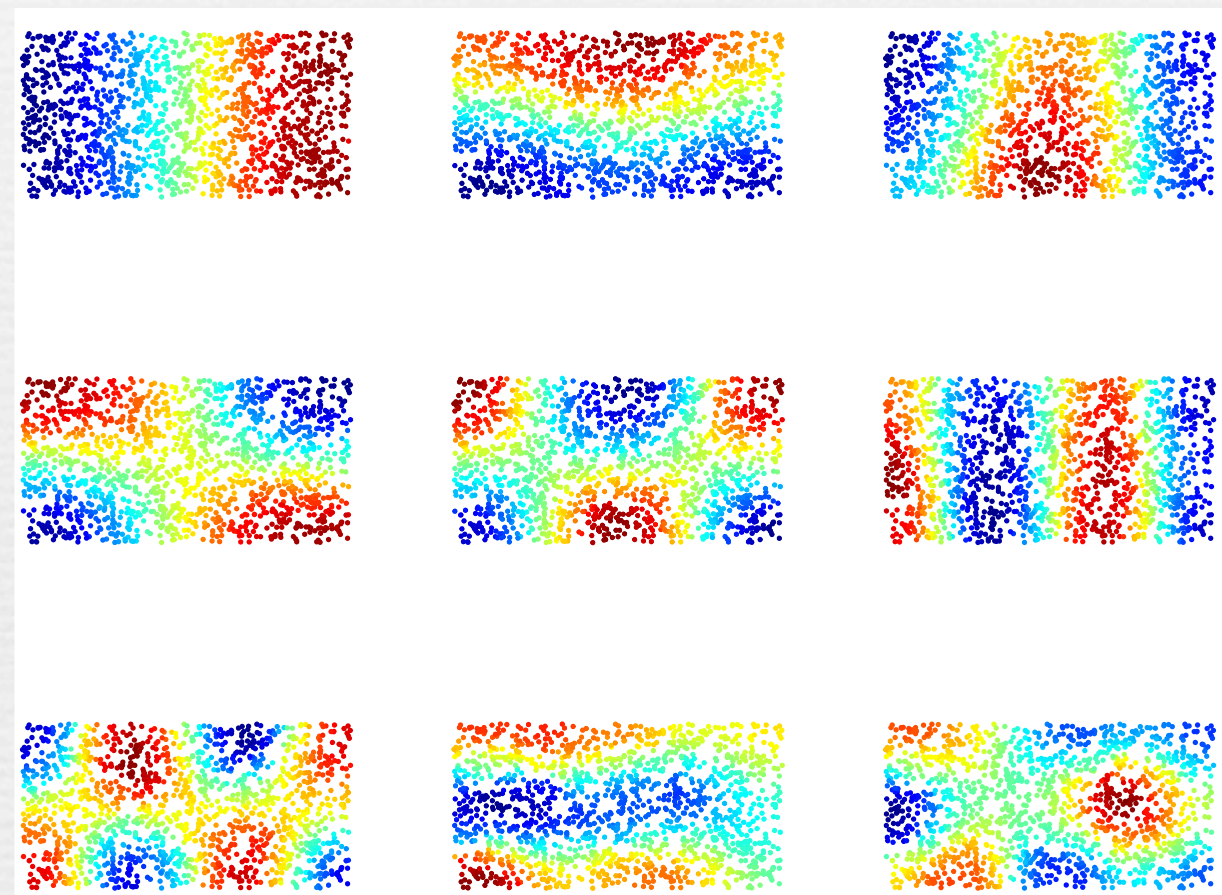
  - eigenfunctions f_1, f_2, f_3, ... as NLDR coordinates

Laplacian Eigenmaps: Swiss Roll

eigenfunctions 1 and 2



eigenfunctions 1 to 9



- eigenfunctions constitute an orthonormal basis for all functions $V \to \mathbf{R}$
- f_0, f_1, f_2, ... successively smoothest functions

NLDR techniques

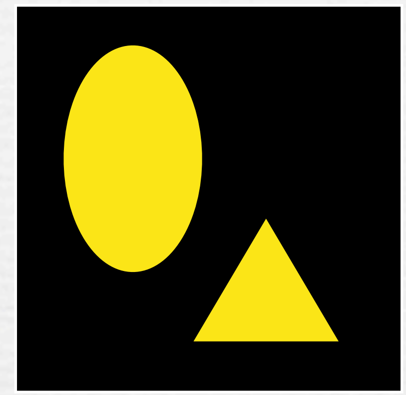
- Since December 2000:
  - Isomap (Tenenbaum, dS, Langford) — geodesics
  - LLE (Roweis, Saul) — local affine structure
  - Laplacian Eigenmaps (Belkin, Niyogi) — diffusion geometry
  - Hessian Eigenmaps (Donoho, Grimes) — 2nd fundamental form
  - ...
- Goal: find useful real-valued coordinate functions on data
  - Most effective when data lie on the image of a convex region
  - Nontrivial topology typically causes problems

What about circle-valued coordinates? $\theta: X \to S^1$

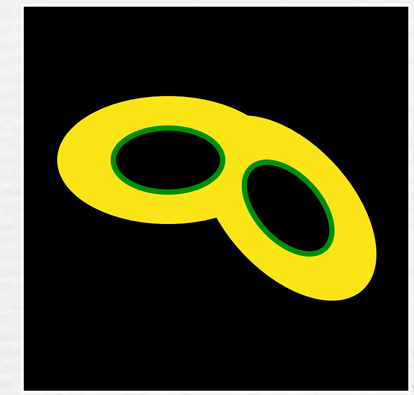
Persistent topology

Betti numbers

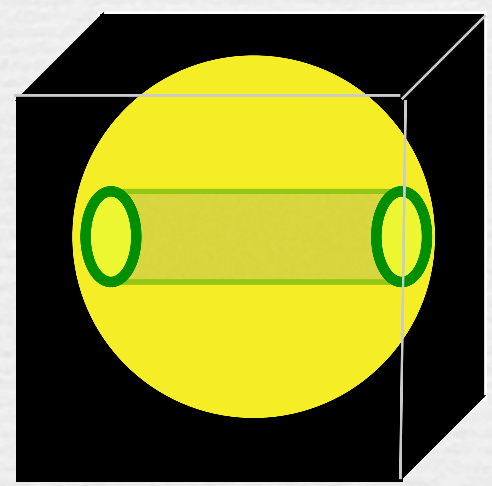
- Objects in 2 dimensions:
    - b_0 = number of components
    - b_1 = number of holes
- Objects in 3 dimensions:
    - b_0 = number of components
    - b_1 = number of tunnels/handles
    - b_2 = number of voids
- (and so on...)



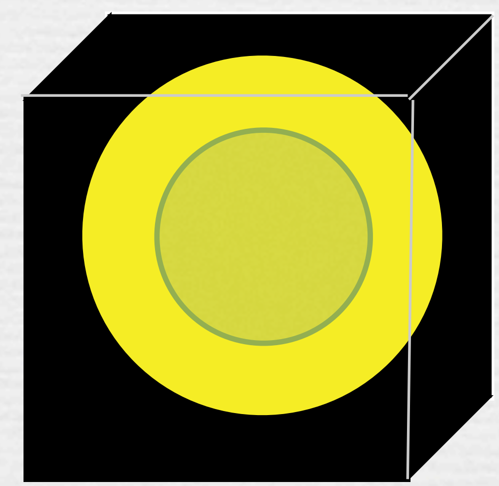
$b_0 = 2, b_1 = 0$



$b_0 = 1, b_1 = 2$

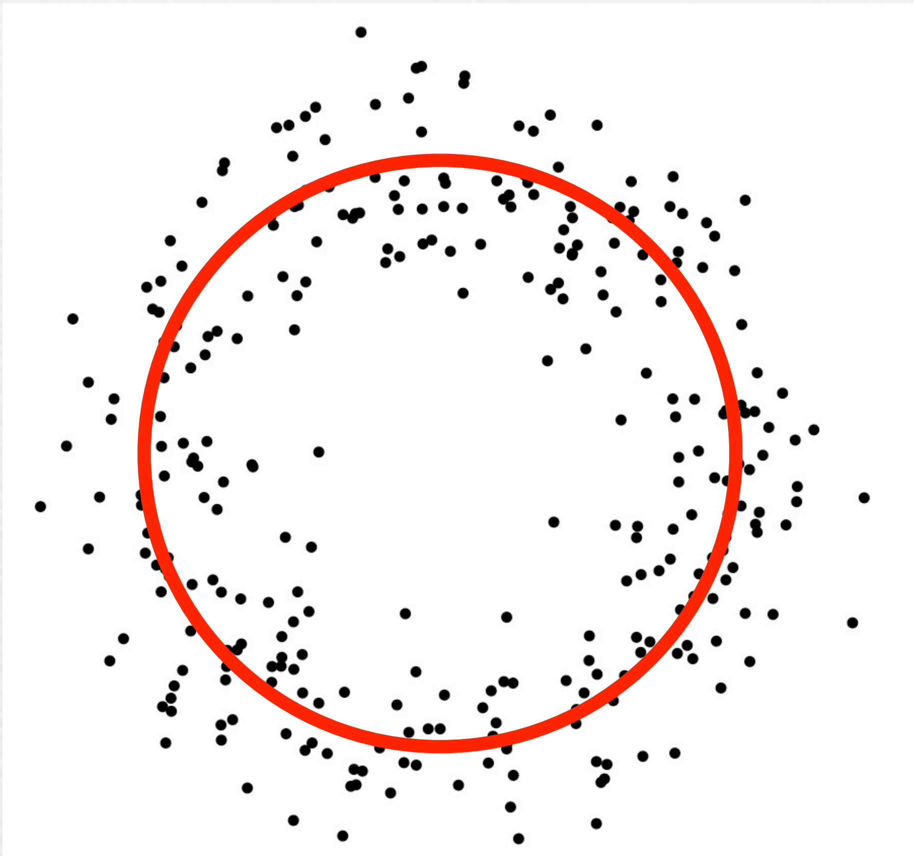


$b_0 = 1, b_1 = 1, b_2 = 0$

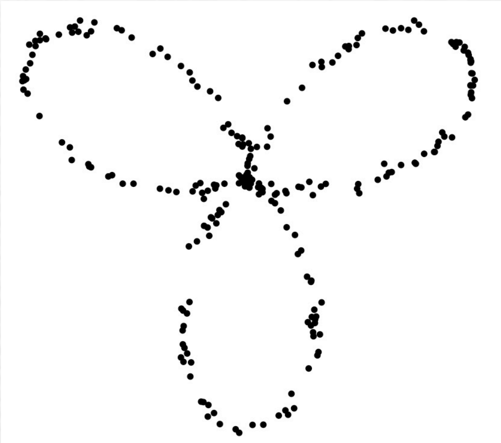


$b_0 = 1, b_1 = 0, b_2 = 1$

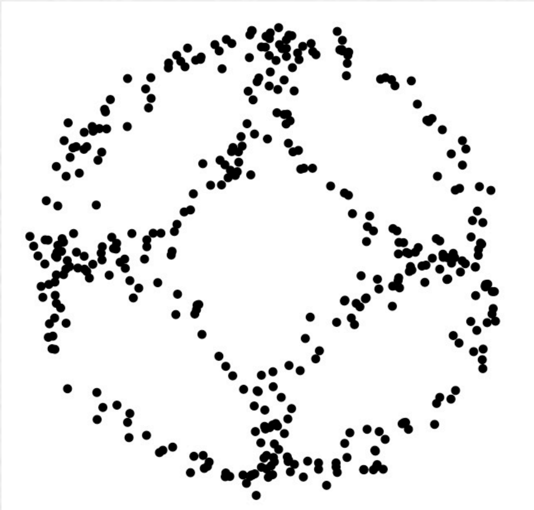
Point-cloud topology



$b_1 = 1$



$b_1 = 3$

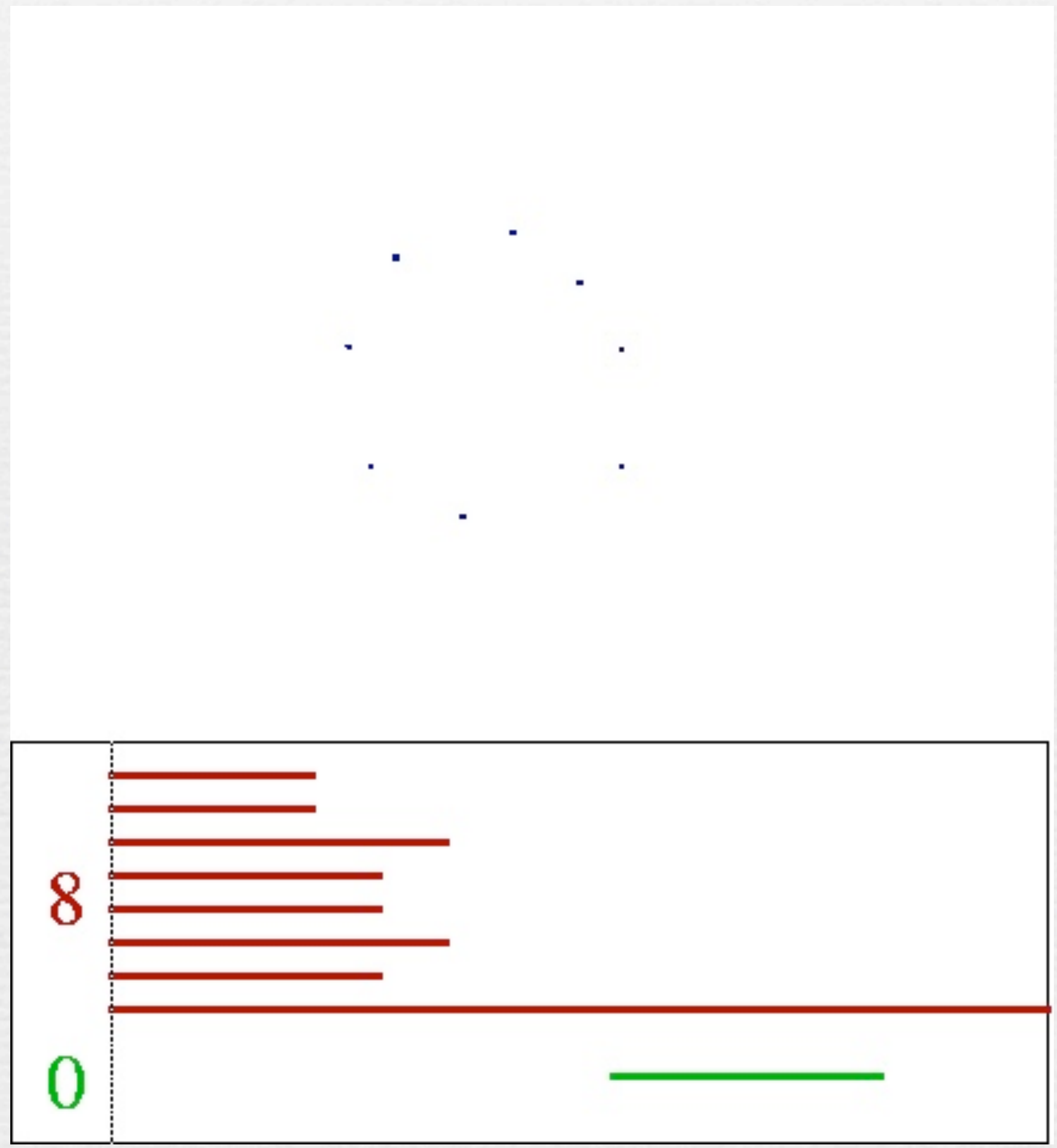


$b_1 = 5$

Data sampled from an unknown topological space Y.
Estimate Betti numbers of Y from the sample.

Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)

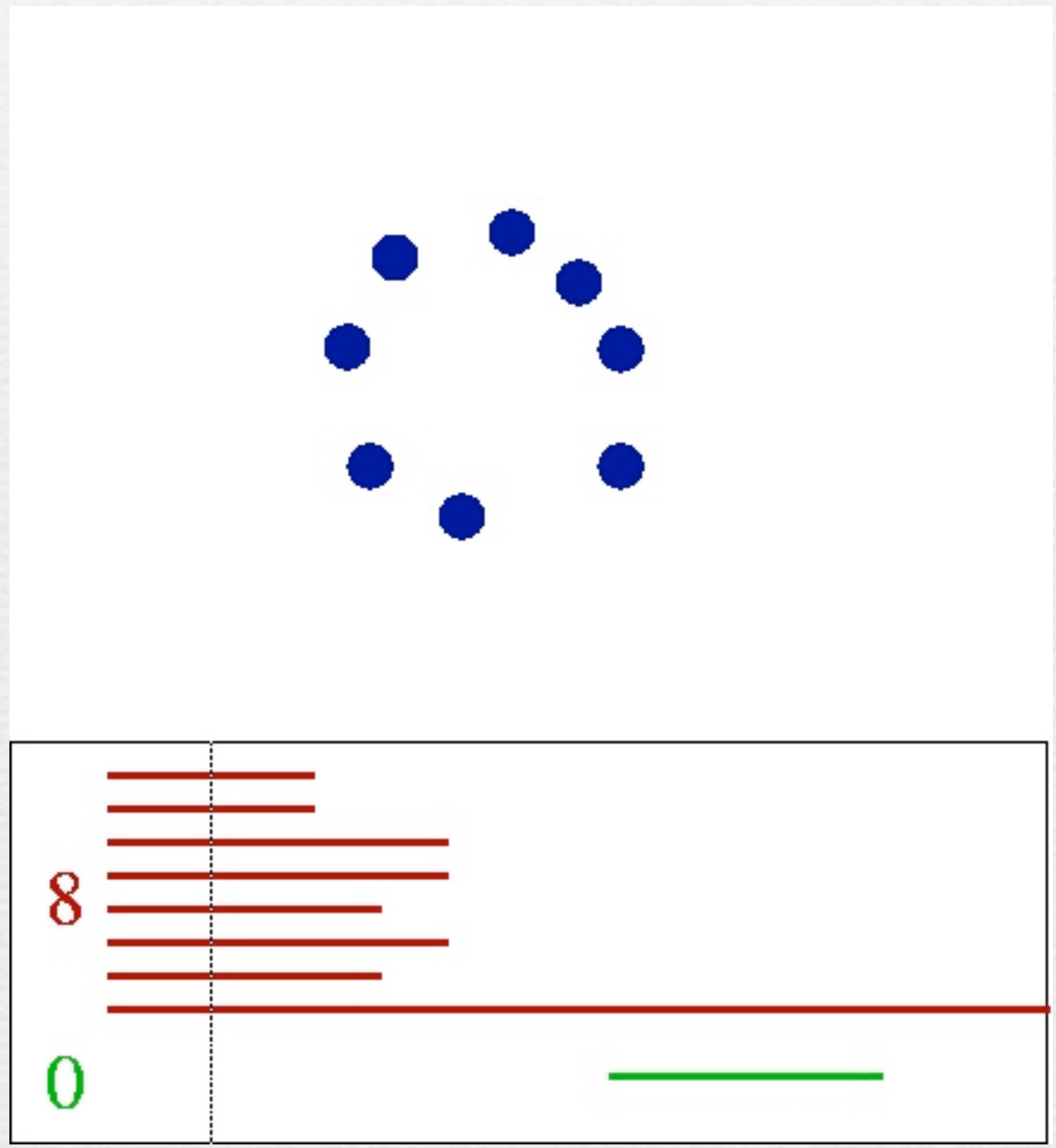


b_0: (connected components)

b_1: (holes)

Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)

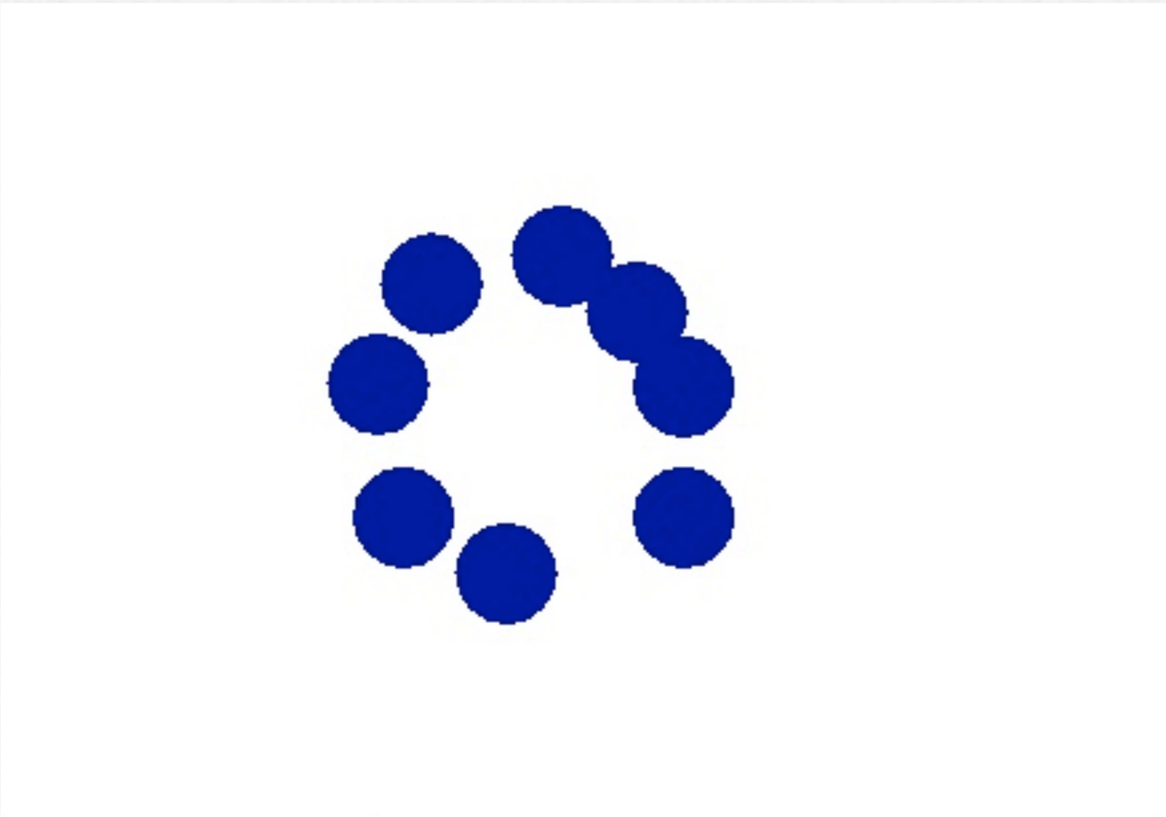


b_0: (connected components)

b_1: (holes)

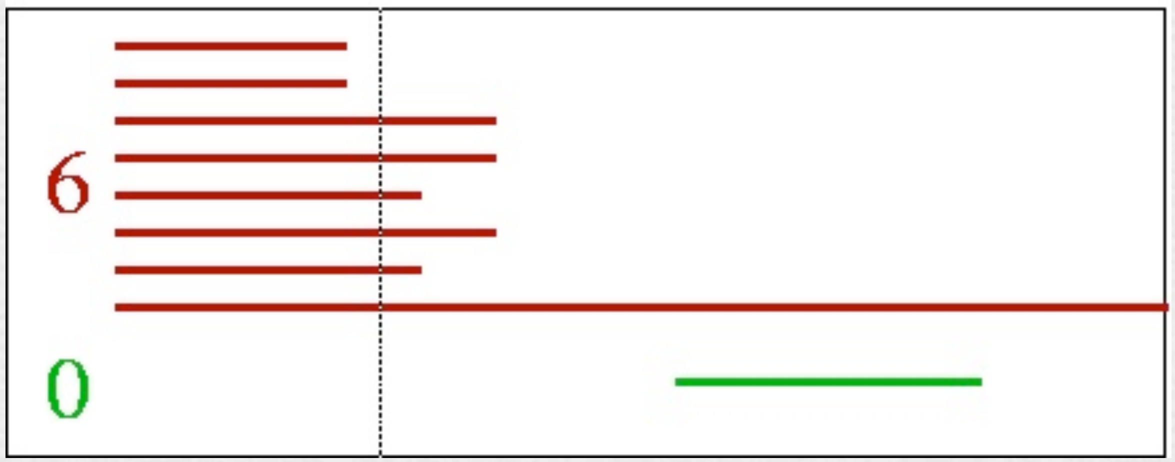
Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)



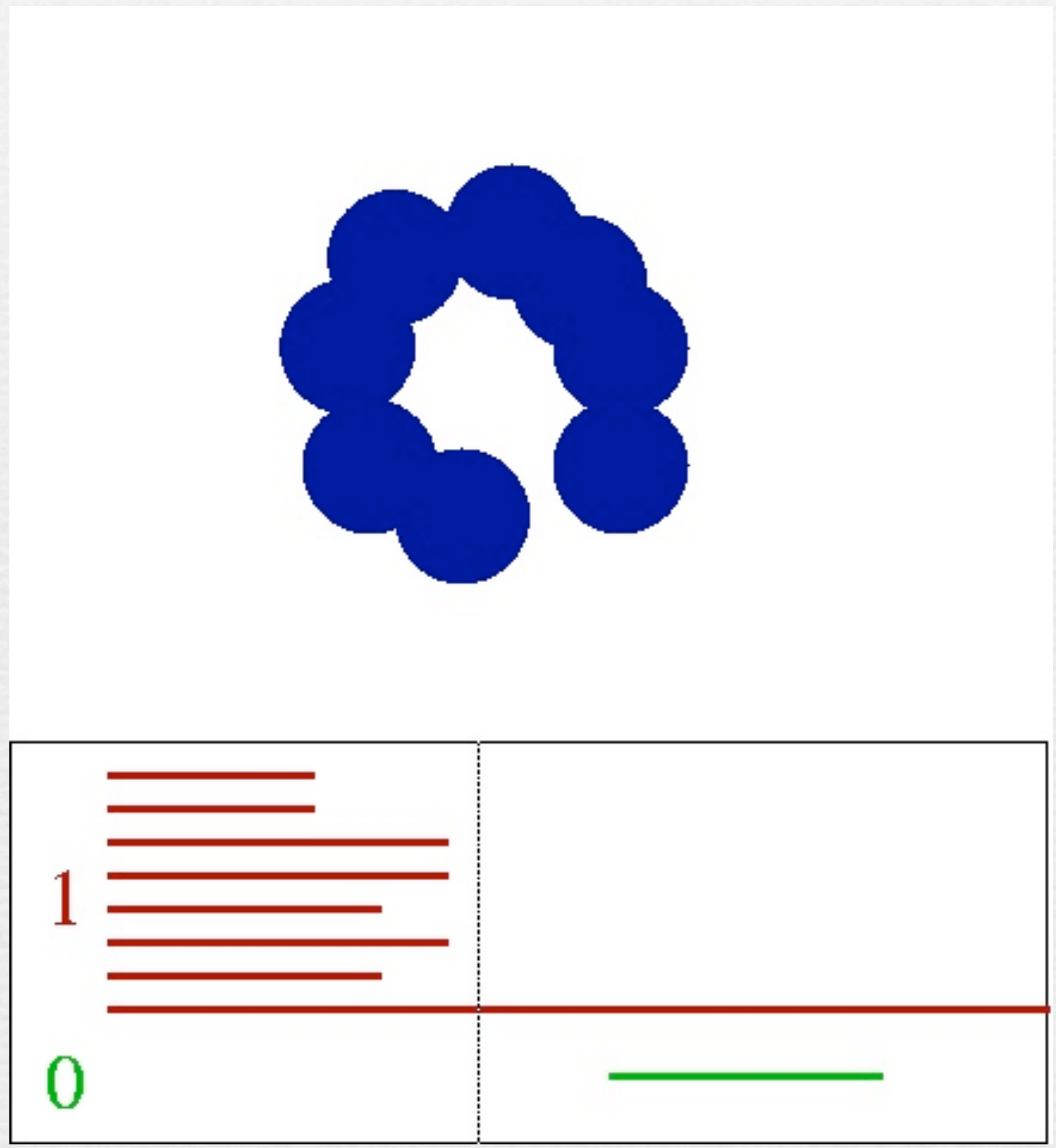
b_0: (connected components)

b_1: (holes)



Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)



b_0: (connected components)

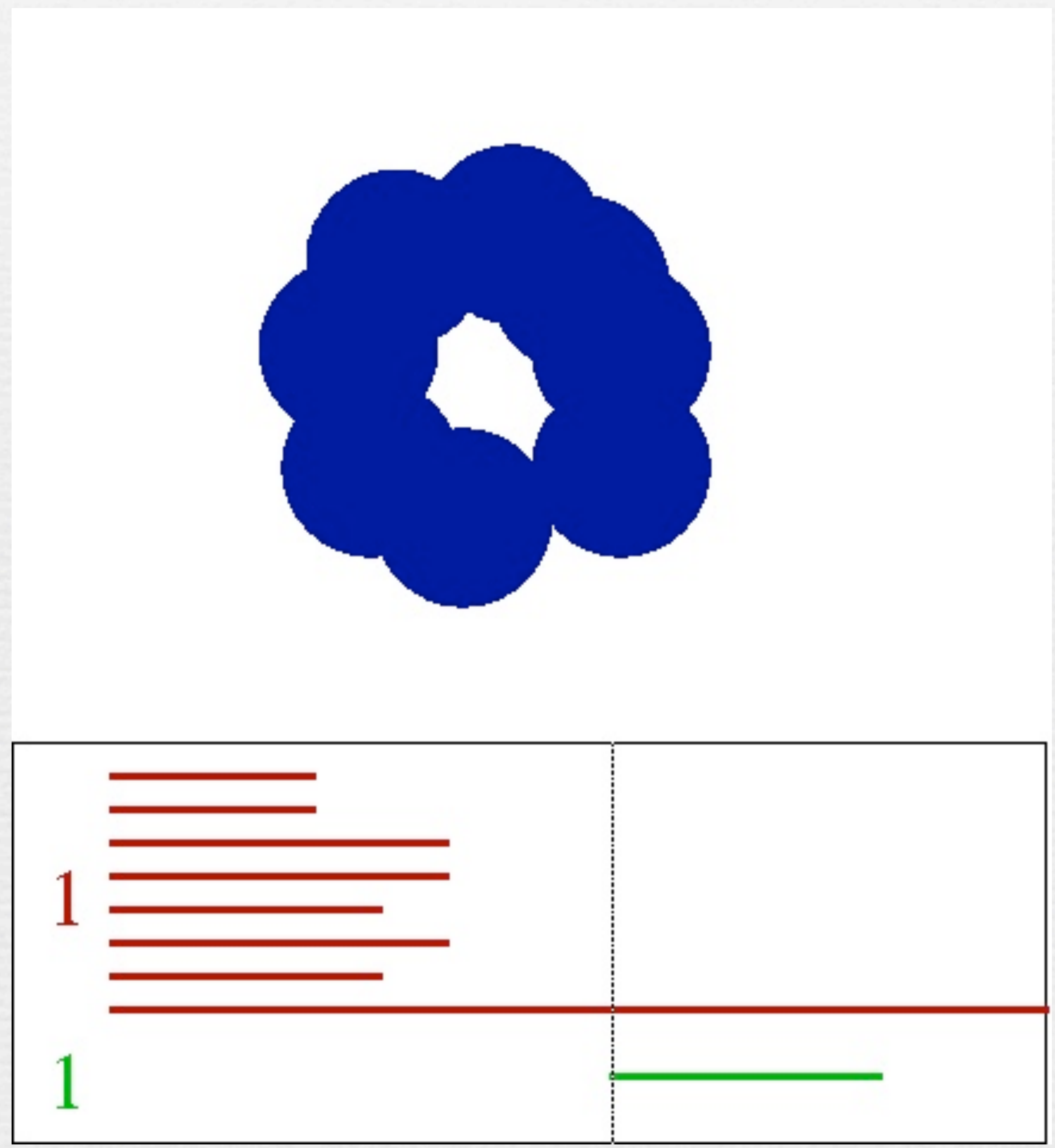
b_1: (holes)

Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)

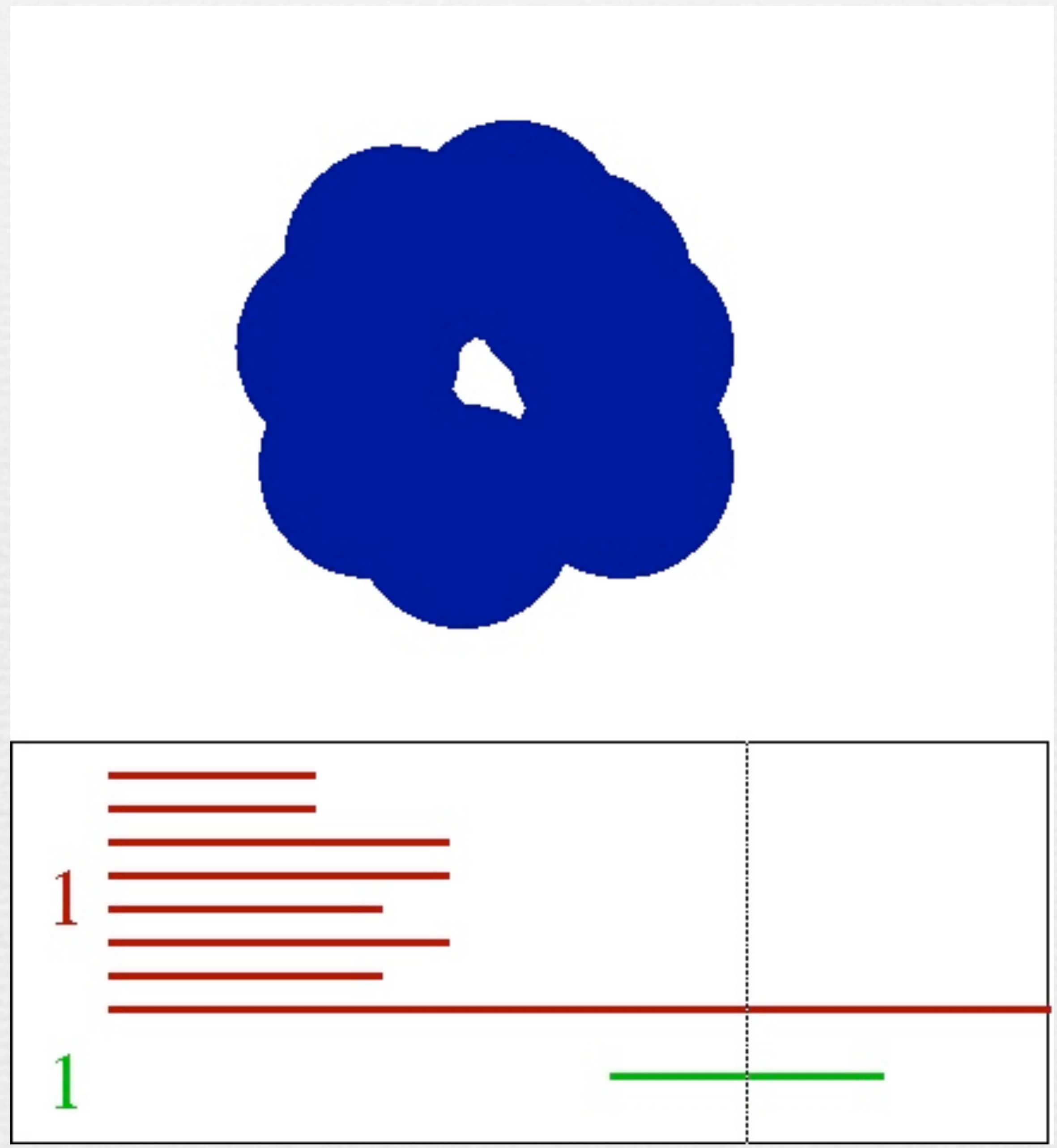
b_0: (connected components)

b_1: (holes)



Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)

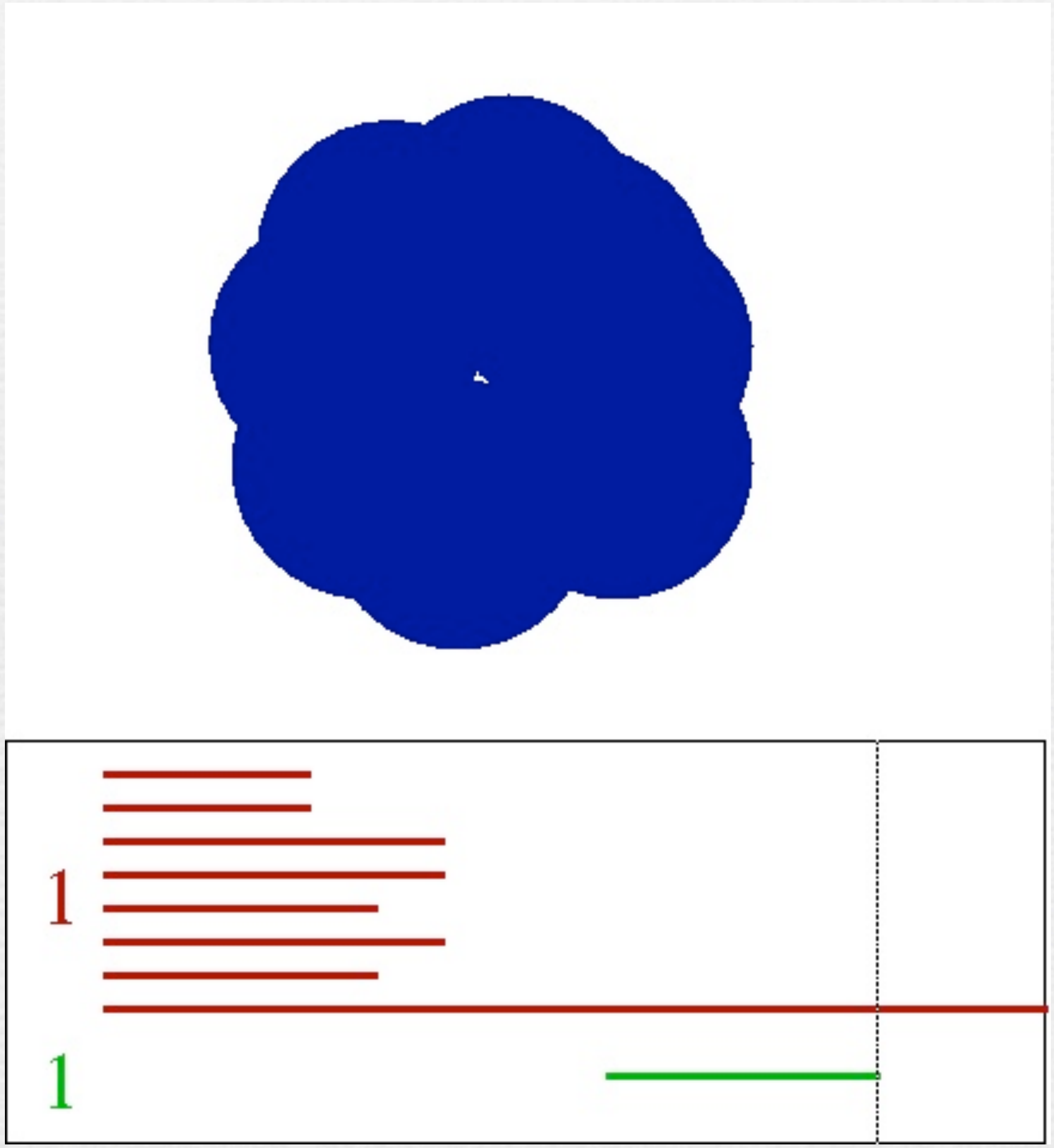


b_0: (connected components)

b_1: (holes)

Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)

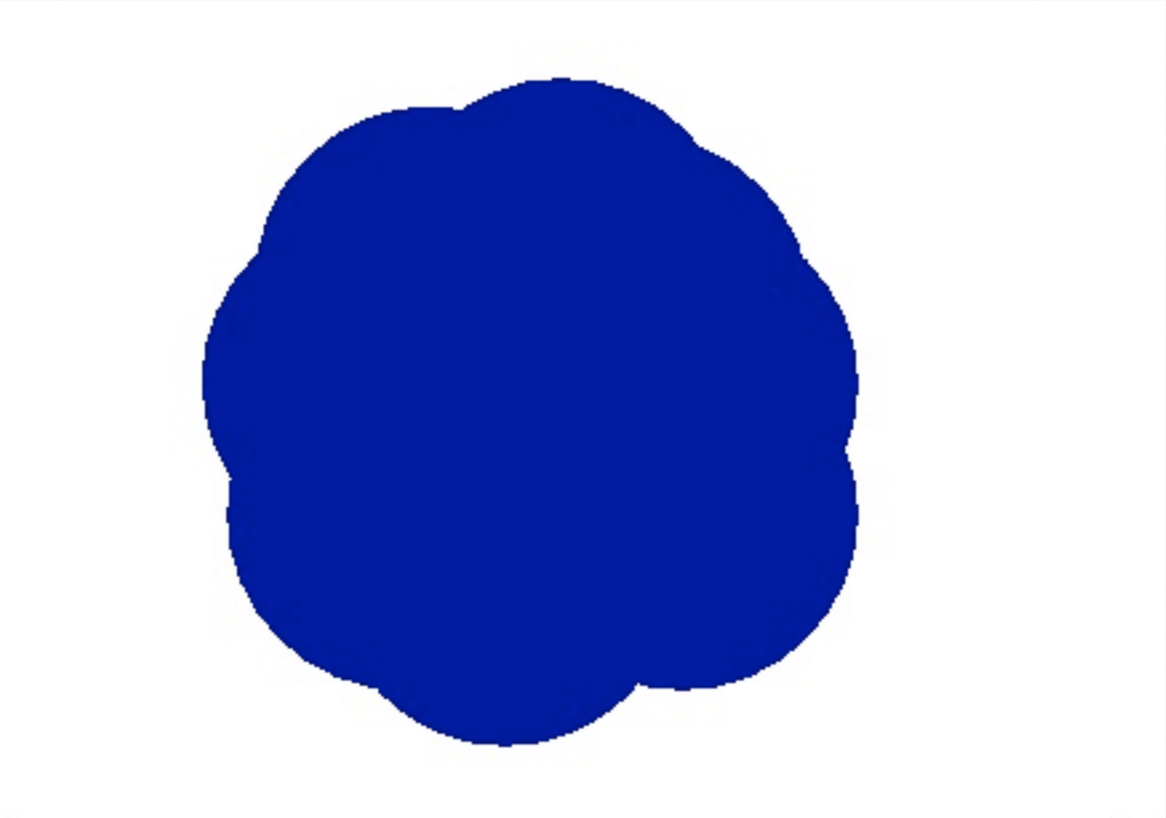


b_0: (connected components)

b_1: (holes)

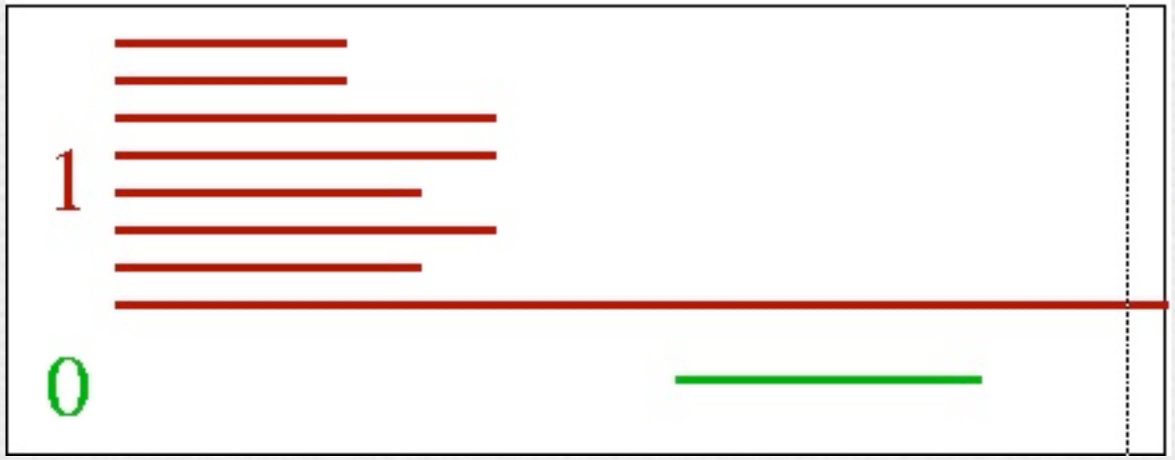
Persistence: robust multiscale topology

Edelsbrunner, Letscher Zomorodian (2000)



b_0: (connected components)

b_1: (holes)



Cohomology

Cohomology

- Represent data by graph, then:
  - cochain spaces

$$C^0 = \text{vector space spanned by vertices} \cong \{f : V \to \mathbb{R}\}$$ scalar fields

$$C^1 = \text{vector space spanned by edges} \cong \{\alpha : E \to \mathbb{R}\}$$ vector fields

$$C^2 = \text{vector space spanned by triangles} \cong \{\alpha : T \to \mathbb{R}\}$$ skew tensor fields

  - coboundary map

$$\delta : C^0 \to C^1; \quad \delta f([ab]) = f(b) - f(a)$$ discrete gradient

$$\delta : C^1 \to C^2; \quad \delta\alpha([abc]) = \alpha([bc]) - \alpha([ac]) + \alpha([ab])$$ discrete curl

  - cohomology

$$\mathsf{H}^0 = \frac{\text{0-cocycles}}{\text{0-coboundaries}} = \frac{\mathrm{Ker}(\delta : C^0 \to C^1)}{0}$$ locally constant scalar fields

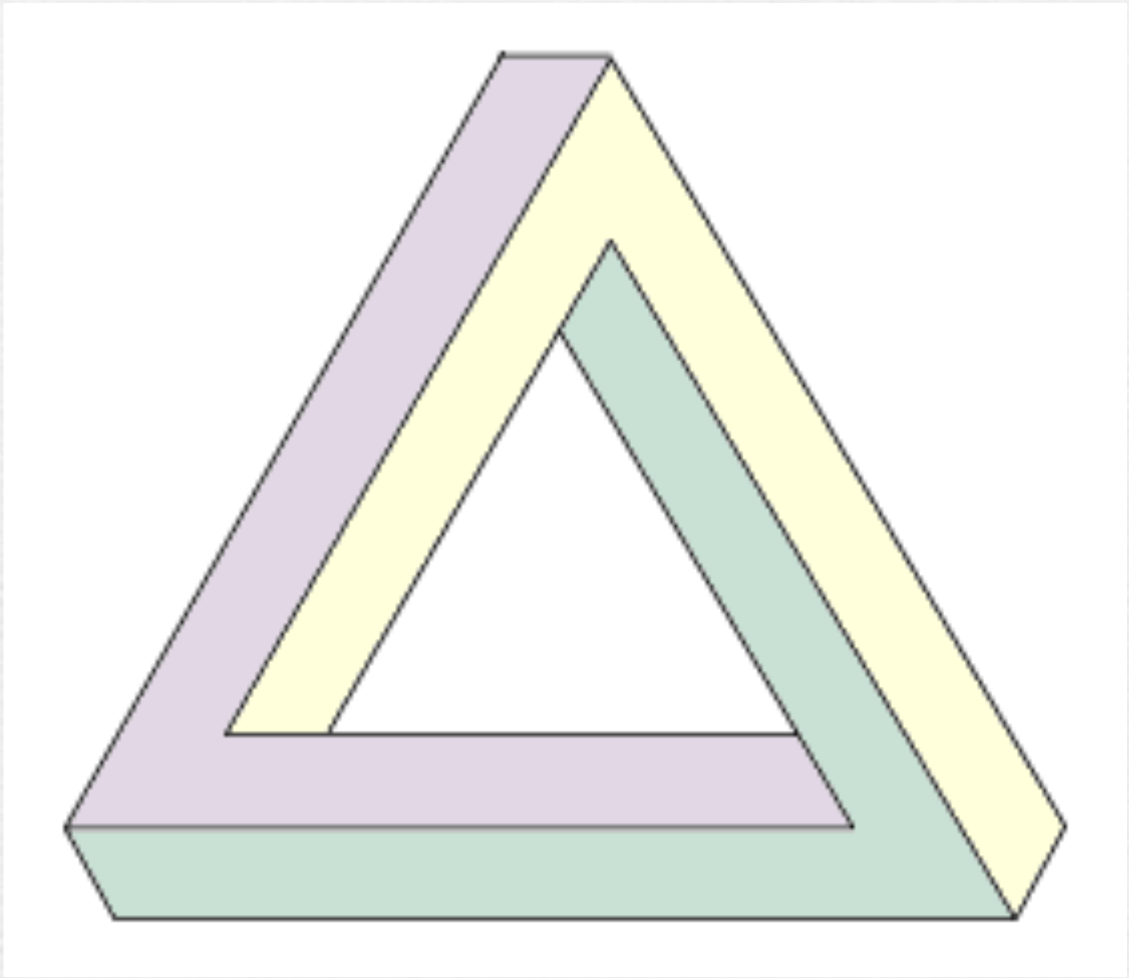
$$\mathsf{H}^1 = \frac{\text{1-cocycles}}{\text{1-coboundaries}} = \frac{\mathrm{Ker}(\delta : C^1 \to C^2)}{\mathrm{Im}(\delta : C^0 \to C^1)}$$ curl-free fields / gradient fields

  - Betti numbers

$b_0 = \dim(H^0)$ number of connected components

$b_1 = \dim(H^1)$ number of 1-dimensional holes

An idea of Roger Penrose...



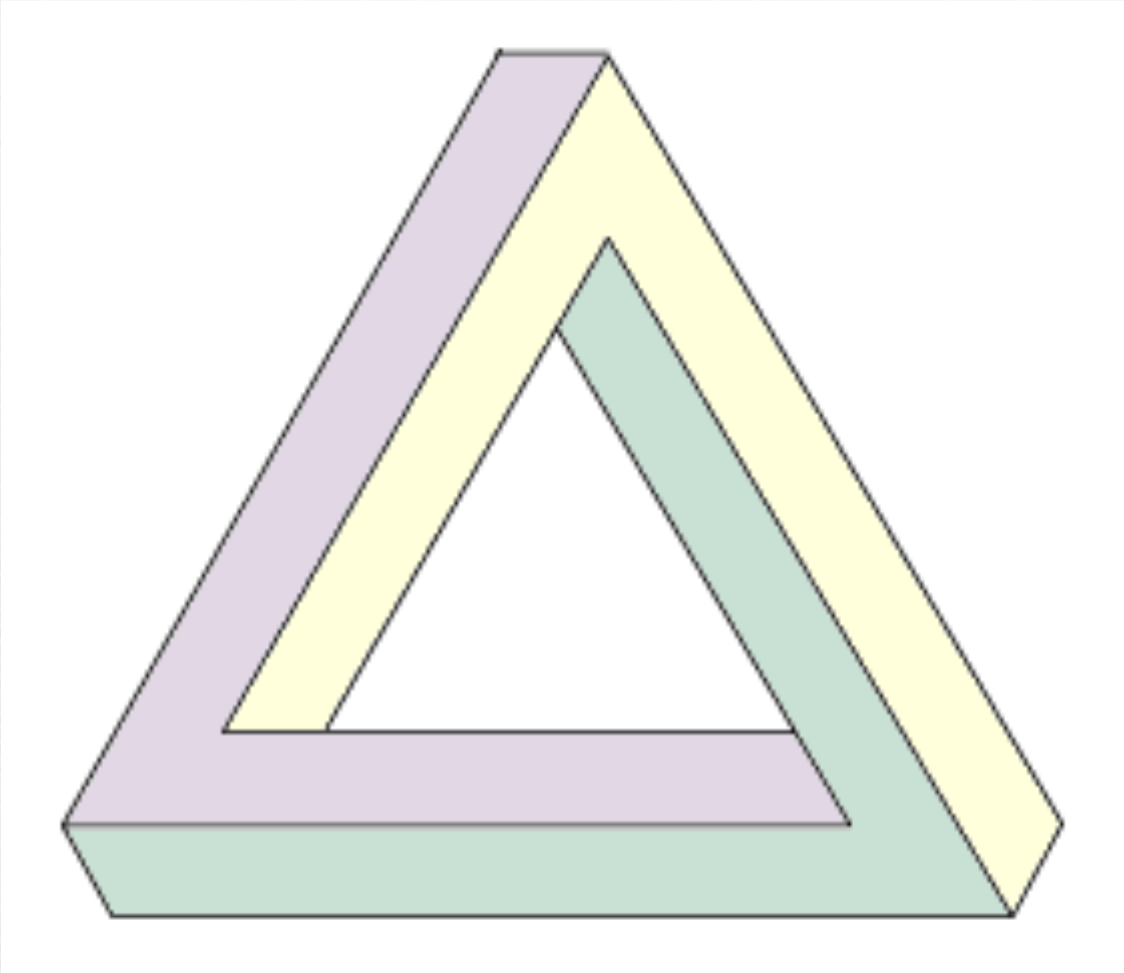
An idea of Roger Penrose...

- What is the depth f(x)?
- g(x,y) = "f(x)-f(y)" is locally consistently defined.

  cohomology class in $H^1(X)$
- There is no global f(x).

  cohomology class is nonzero
- f(x) is definable modulo integral around triangle.

  circle-valued depth function



cohomology $H^1(X)$ = locally consistent g(x,y) / globally consistent f(x)-f(y)

Circular coordinates (dS, Morozov, Vejdemo-Johansson)

- Classical equation from homotopy theory:

$$[X, S^1] = H^1(X; \mathbf{Z})$$

Integer cohomology of X

Homotopy classes of maps $X \to S^1$

- To find circular coordinates:
  - find integer 1-cocycles of high robustness
  - project onto the kernel of the 1-Laplacian (for smoothness)
  - integrate the 1-cocycles to functions onto **R/Z**

Interpretation via graph flows

- Oriented flow (on edges)

  $\alpha : \mathrm{Edges}(X) \to \mathbb{Z}$

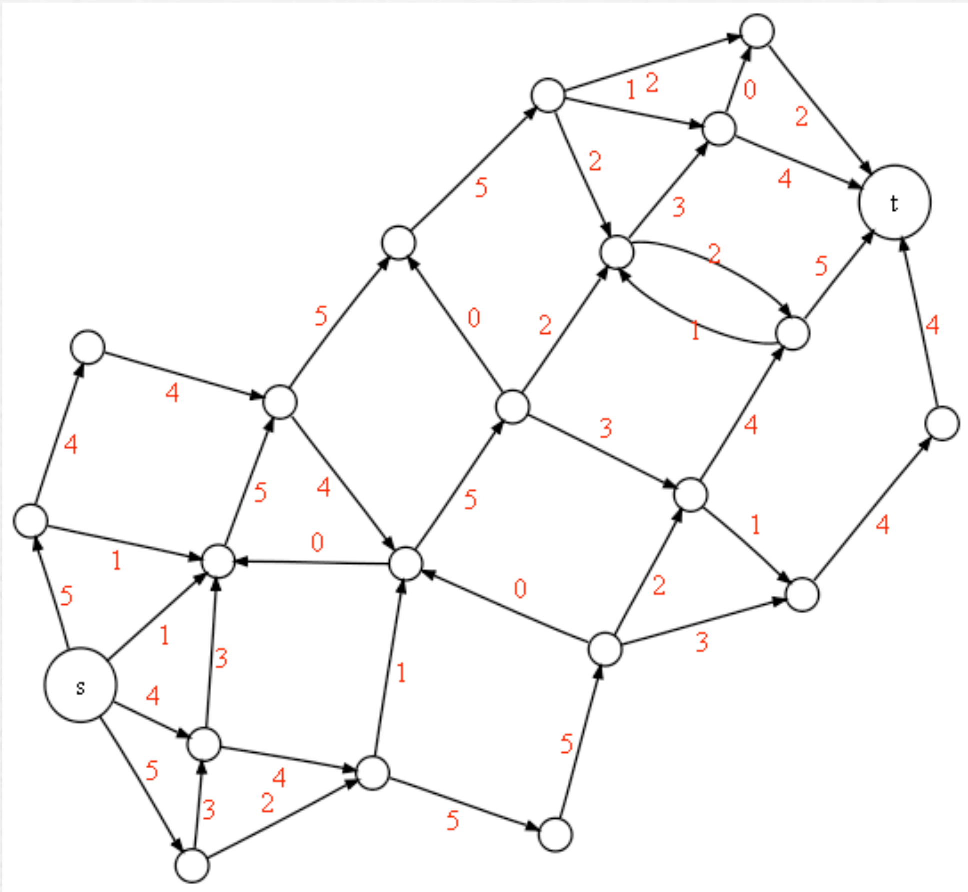
  $\alpha : \mathrm{Edges}(X) \to \mathbb{R}$

- Cocycle condition

  Net flow around each triangle is zero

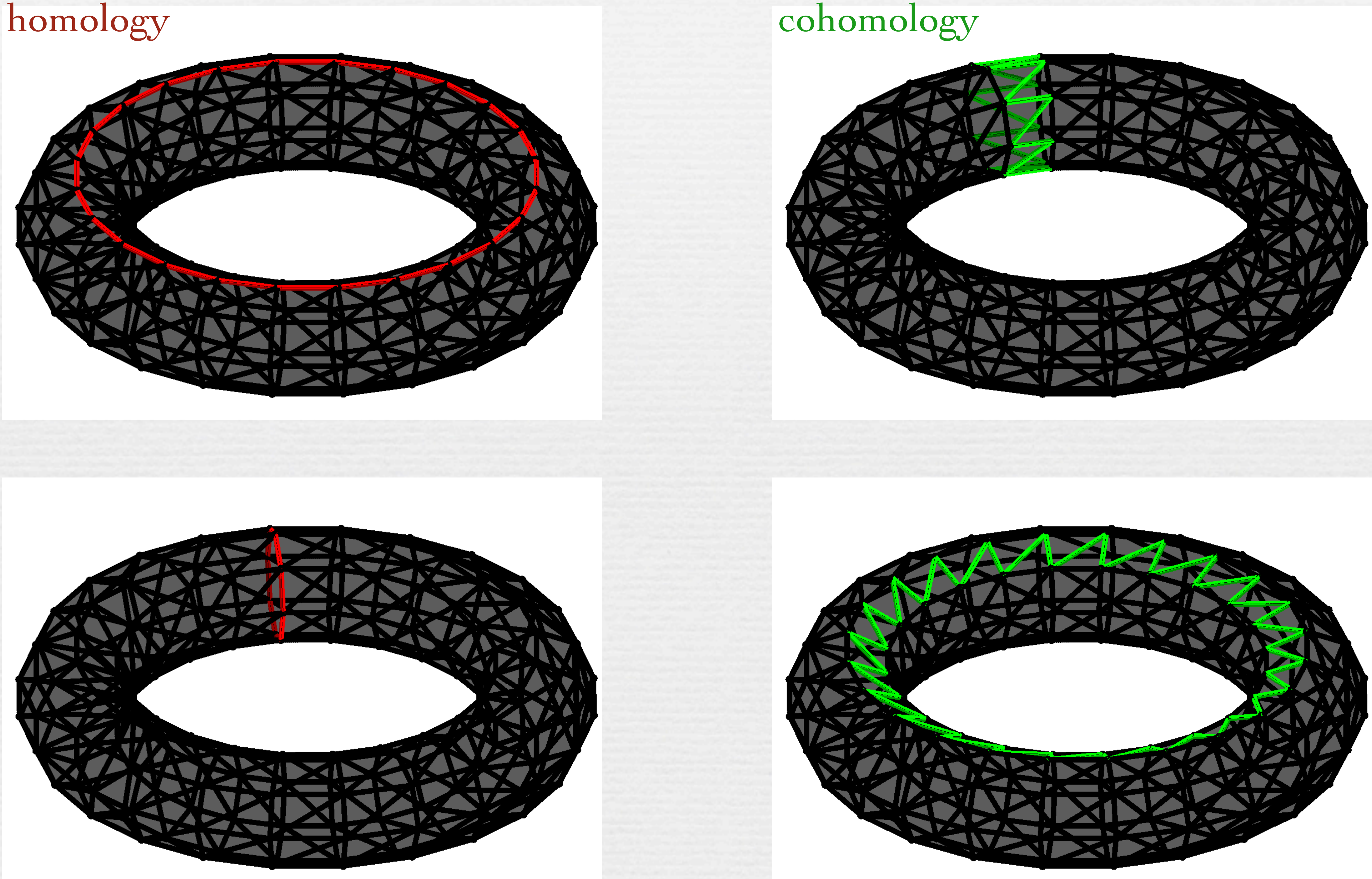
- Cycle condition

  Net flow into each vertex is zero



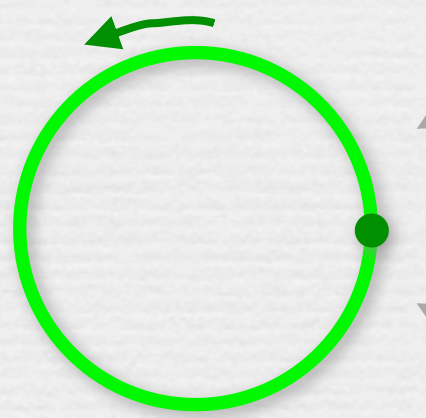
Find an integer flow satisfying the cocycle condition.
Smoothe to a real flow by imposing the cycle condition (L^2-nearest).

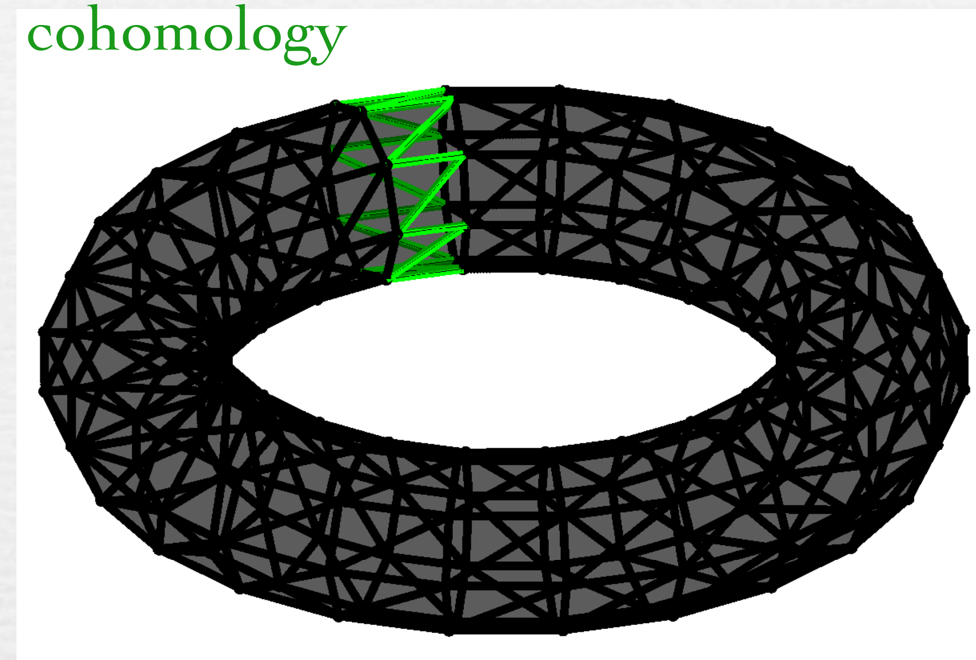
Dual bases

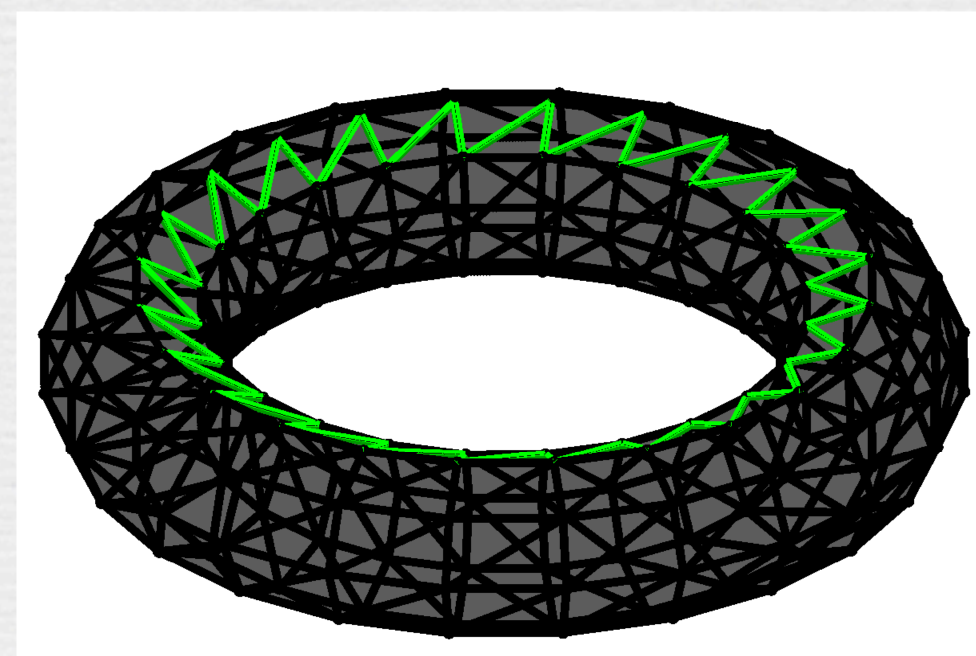


Co-circles

- Integer cocycle a gives rise to circle map:
  - vertices map to base point
  - edge ab winds k times around circle, where k = a(ab)

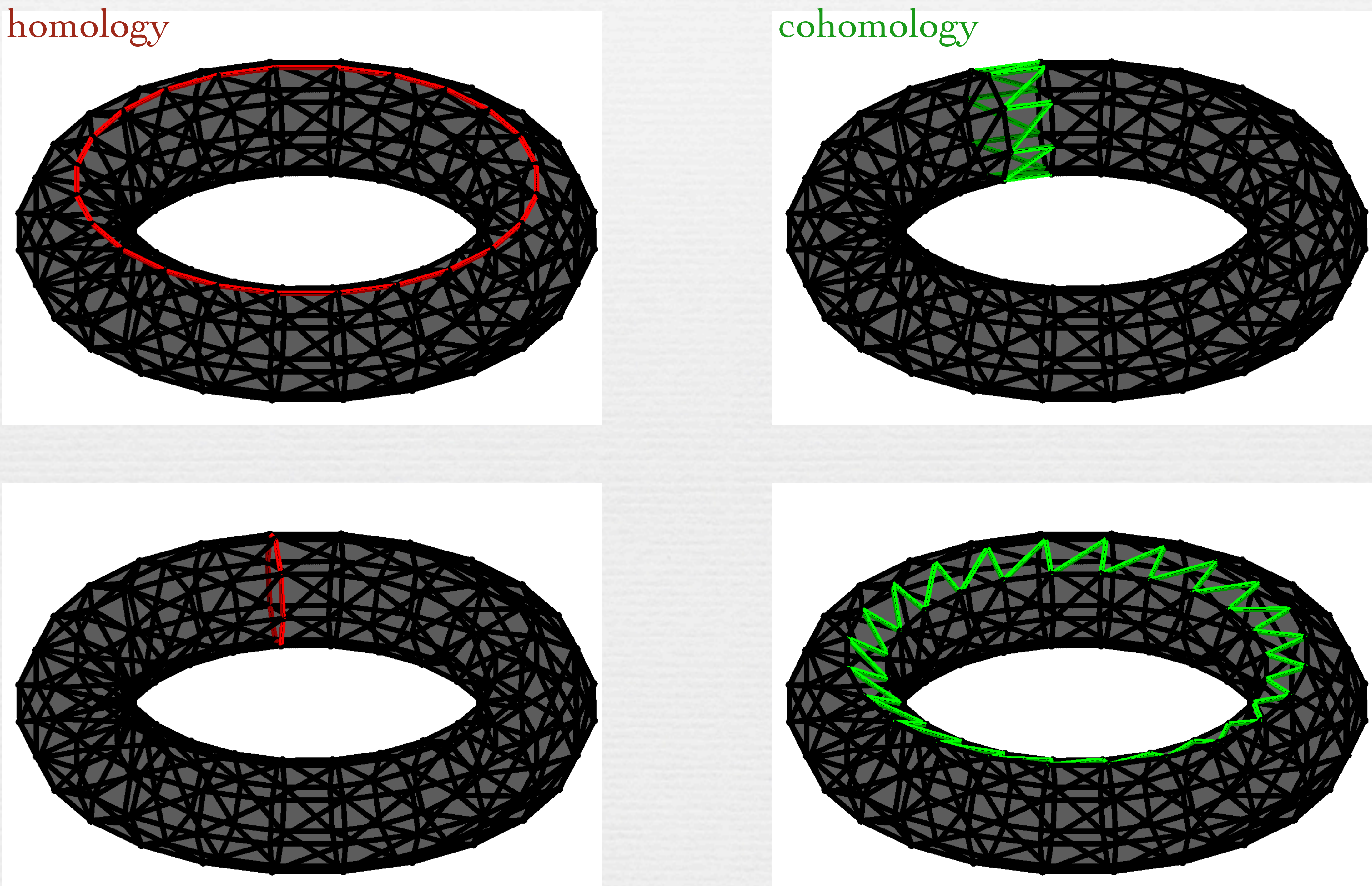




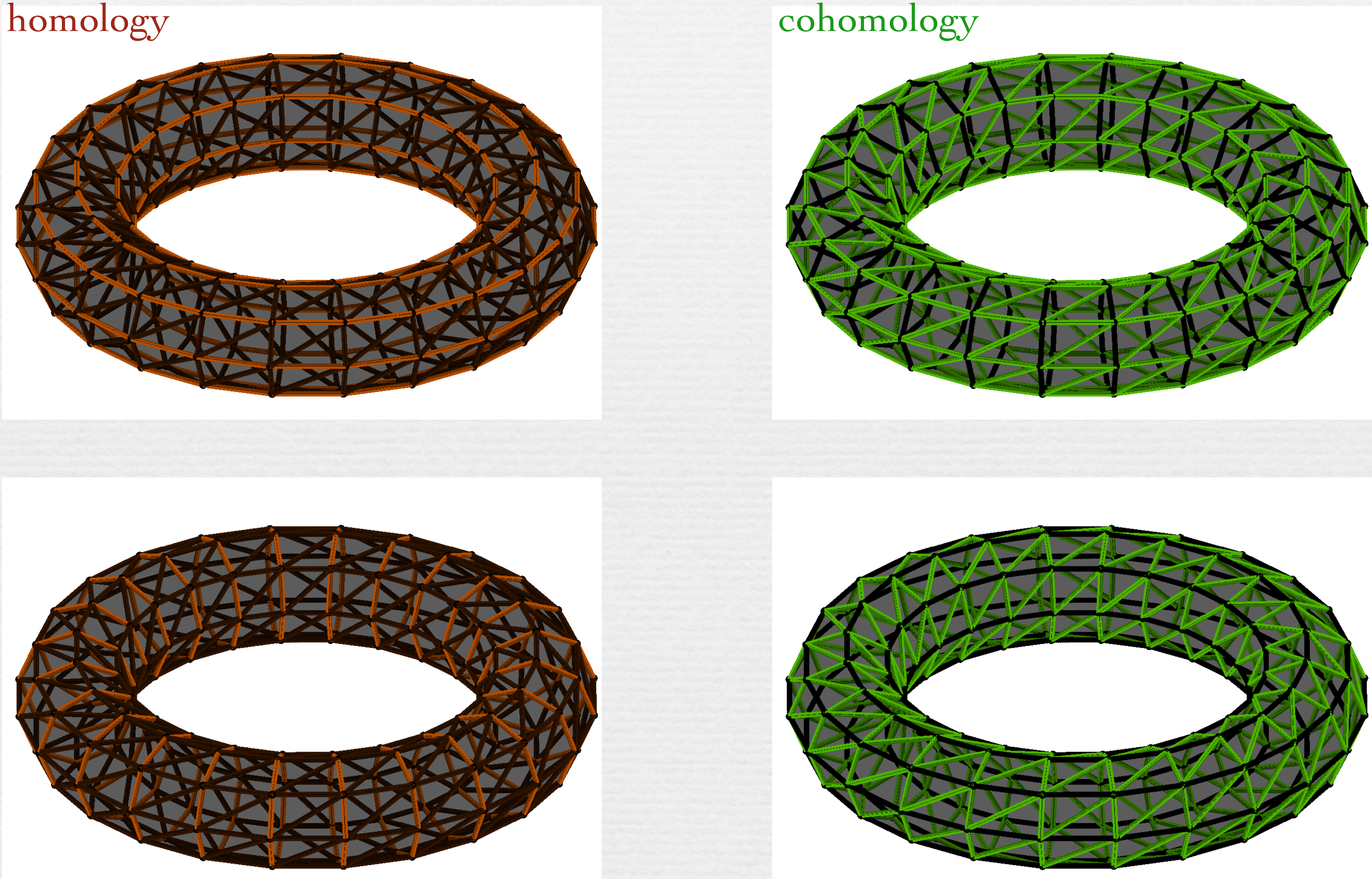


  - cocycle condition guarantees that the map can be extended over triangles
- Not very smooth.

Harmonic smoothing

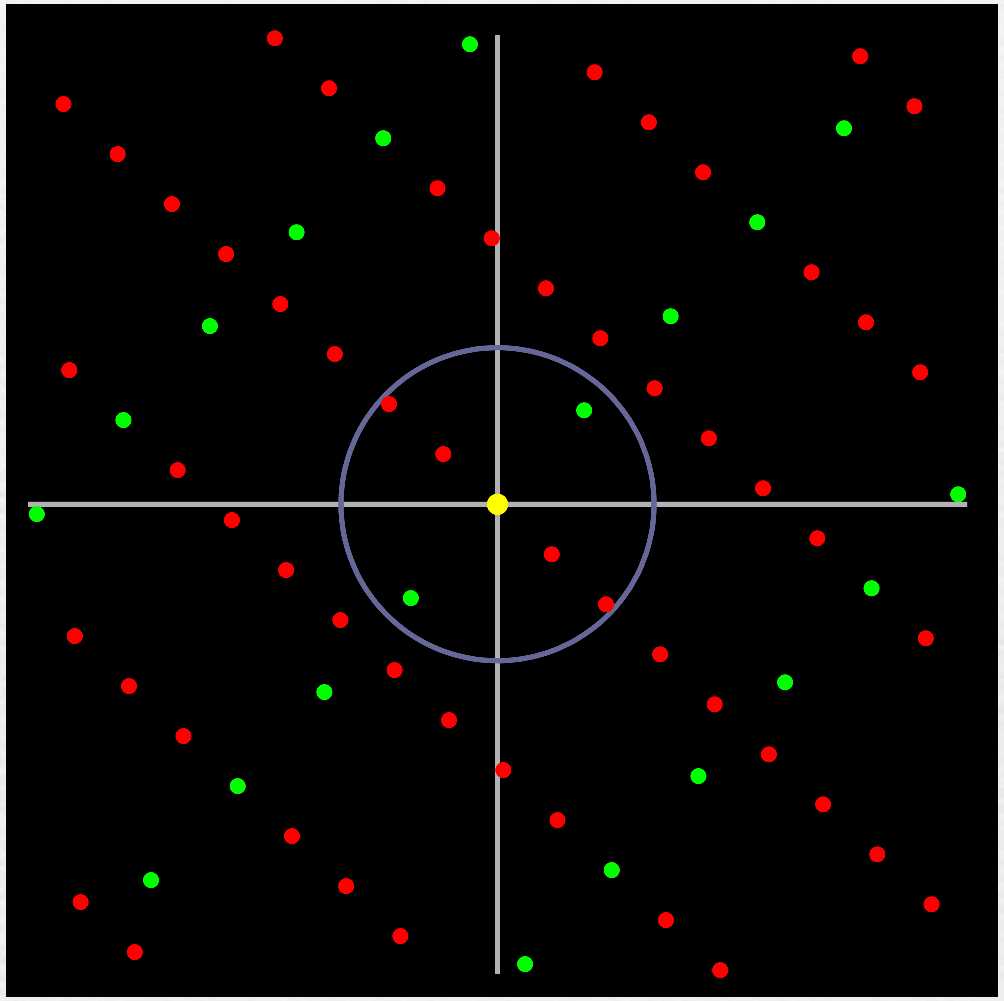


Harmonic smoothing



Hodge theory

harmonic 1-forms



= graph flows which satisfy cycle & cocycle conditions:

$$\partial\alpha([v]) = 0 \quad \text{for all } v \in \text{Vertices}(X)$$

$$\delta\alpha([uvw]) = 0 \quad \text{for all } [uvw] \in \text{Triangles}(X)$$

smooth circular coordinates
↕
harmonic forms in the
integer cohomology lattice

integer homology and cohomology lattices

$$\mathrm{H}_1(X;\mathbb{Z}) \to \mathrm{H}_1(X;\mathbb{R}) = \mathcal{H}^1(X) = \mathrm{H}^1(X;\mathbb{R}) \leftarrow \mathrm{H}^1(X;\mathbb{Z})$$

$$\begin{aligned} C^1 &= \text{1-coboundaries} \oplus \mathcal{H}^1 \oplus \text{1-boundaries} \\ &= \mathrm{Im}(\delta : C^0 \to C^1) \oplus \mathcal{H}^1 \oplus \mathrm{Im}(\partial : C^2 \to C^1) \end{aligned}$$

real-valued functions
(Belkin–Niyogi)

circle-valued functions (in cohomology lattice)

see also: Statistical ranking with Hodge theory (Jiang, Lim, Yao, Ye)

(The Abel–Jacobi map)

- Let H denote the 1-harmonic space of X
- Let L^1 denote the integer cohomology lattice in H
- Let L_1 denote the integer homology lattice in H

$$\boxed{L^1 \xrightarrow{\text{AJ}} \mathrm{Maps}(X, S^1)} \Leftrightarrow \text{AJ} \in \mathrm{Maps}(L^1 \times X, S^1)$$

$$\Leftrightarrow X \xrightarrow{\text{AJ}} \mathrm{Maps}(L^1, S^1)$$

$$\Leftrightarrow X \xrightarrow{\text{AJ}} \left[\frac{\mathrm{Maps}(L^1, \mathbb{R})}{\mathrm{Maps}(L^1, \mathbb{Z})}\right]$$

$$\Leftrightarrow X \xrightarrow{\text{AJ}} H/L_1$$

(The Abel–Jacobi map)

- Let H denote the 1-harmonic space of X
- Let L^1 denote the integer cohomology lattice in H
- Let L_1 denote the integer homology lattice in H

$$L^1 \xrightarrow{\text{AJ}} \mathrm{Maps}(X, S^1) \Leftrightarrow \text{AJ} \in \mathrm{Maps}(L^1 \times X, S^1)$$

$$\Leftrightarrow X \xrightarrow{\text{AJ}} \mathrm{Maps}(L^1, S^1)$$

$$\Leftrightarrow X \xrightarrow{\text{AJ}} \left[\frac{\mathrm{Maps}(L^1, \mathbb{R})}{\mathrm{Maps}(L^1, \mathbb{Z})}\right]$$

$$\Leftrightarrow \boxed{X \xrightarrow{\text{AJ}} H/L_1}$$

(The Abel–Jacobi map)

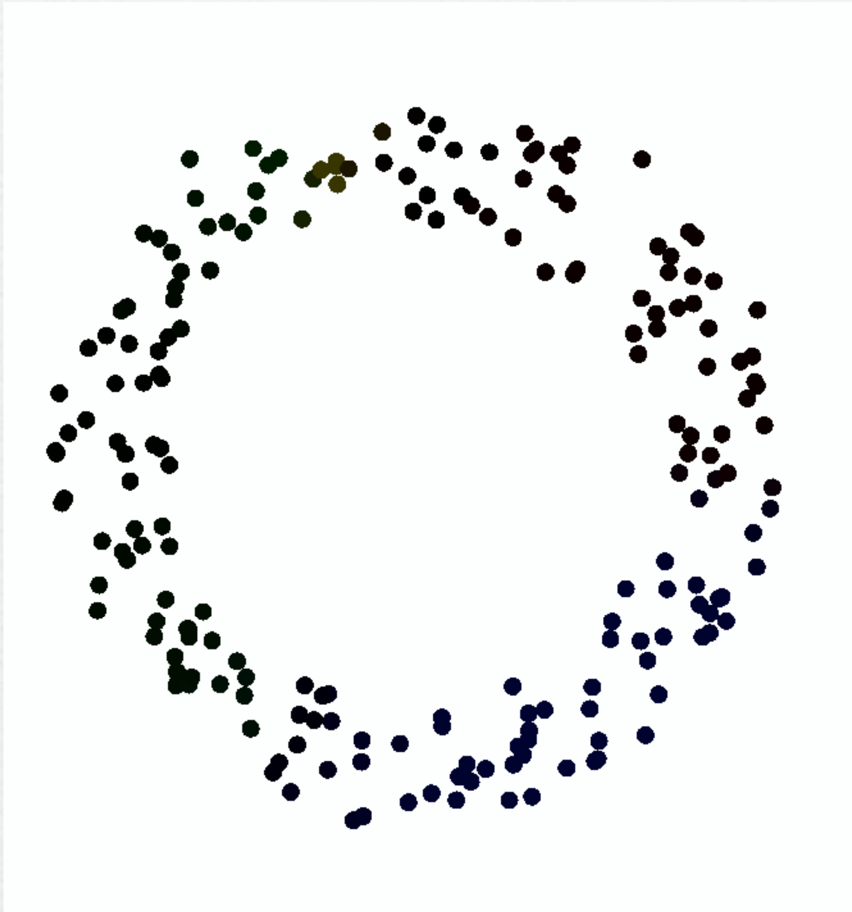
- Let H denote the 1-harmonic space of X
- Let L^1 denote the integer cohomology lattice in H
- Let L_1 denote the integer homology lattice in H

$$
\begin{aligned}
L^1 \xrightarrow{\mathrm{AJ}} \mathrm{Maps}(X, S^1) &\Leftrightarrow \mathrm{AJ} \in \mathrm{Maps}(L^1 \times X, S^1) \\
&\Leftrightarrow X \xrightarrow{\mathrm{AJ}} \mathrm{Maps}(L^1, S^1) \\
&\Leftrightarrow X \xrightarrow{\mathrm{AJ}} \left[\frac{\mathrm{Maps}(L^1, \mathbb{R})}{\mathrm{Maps}(L^1, \mathbb{Z})}\right] \\
&\Leftrightarrow X \xrightarrow{\mathrm{AJ}} H/L_1
\end{aligned}
$$

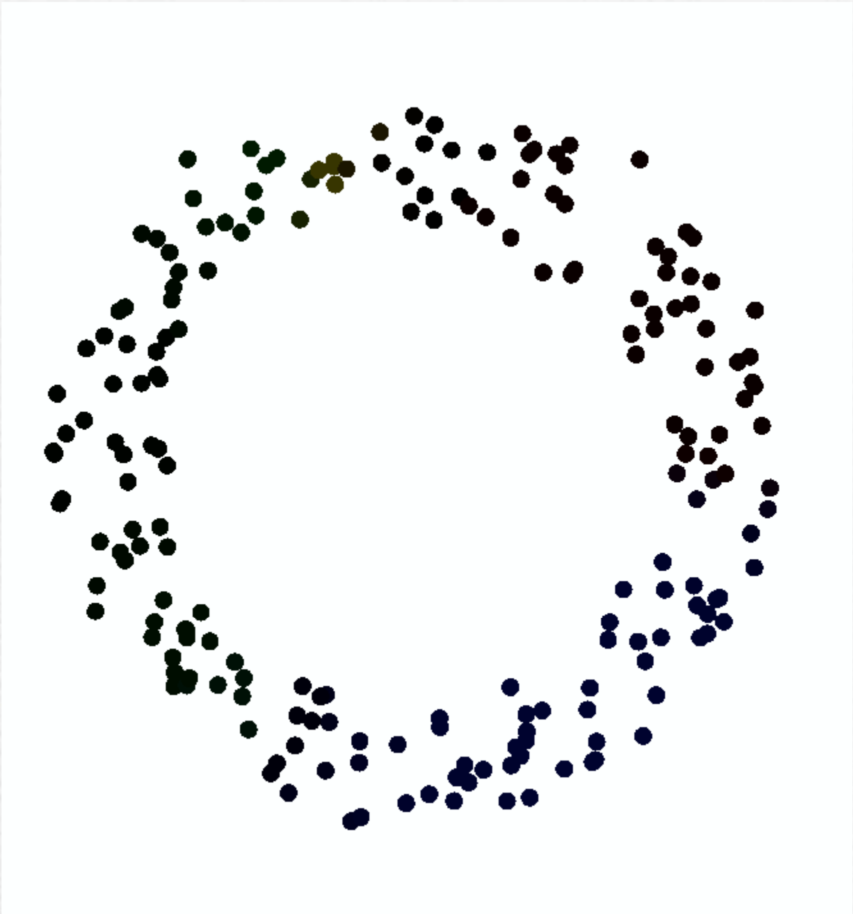
Jacobi torus of X

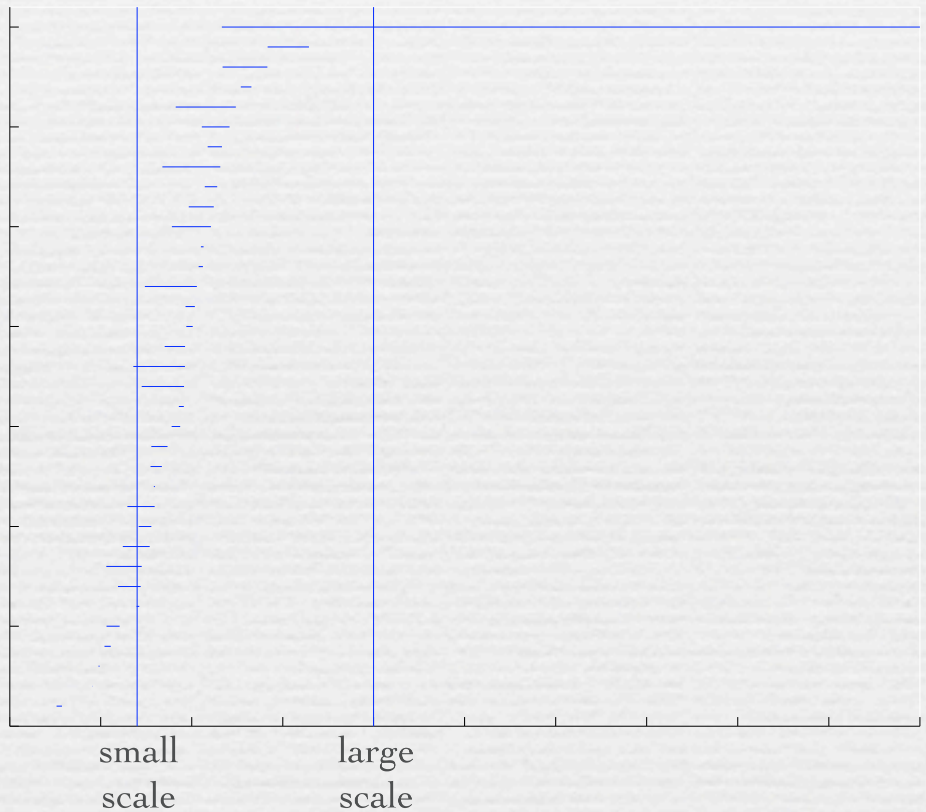
Static data

Noisy circle



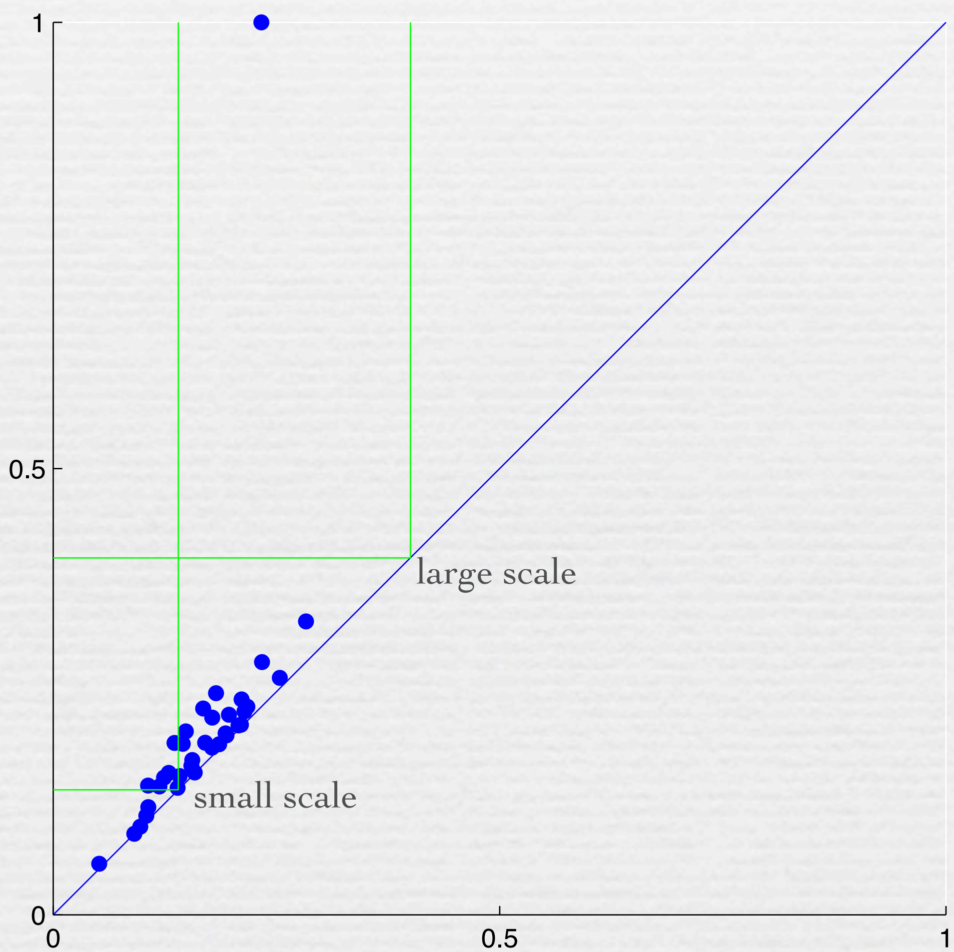
Noisy circle

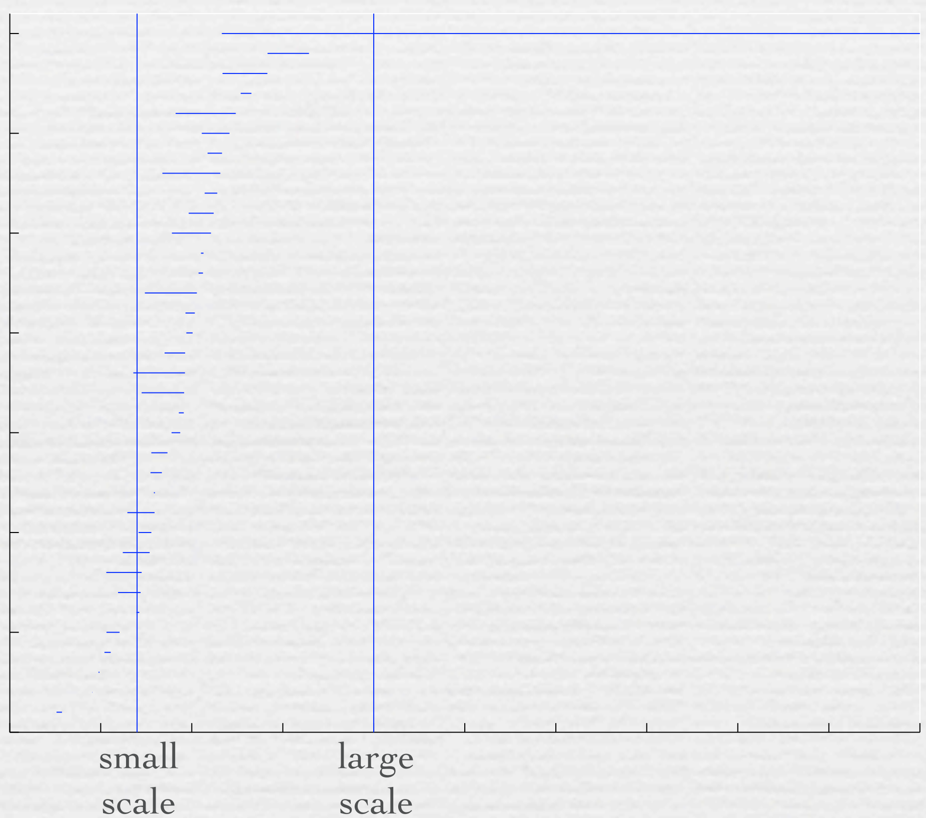




Barcode

Noisy circle





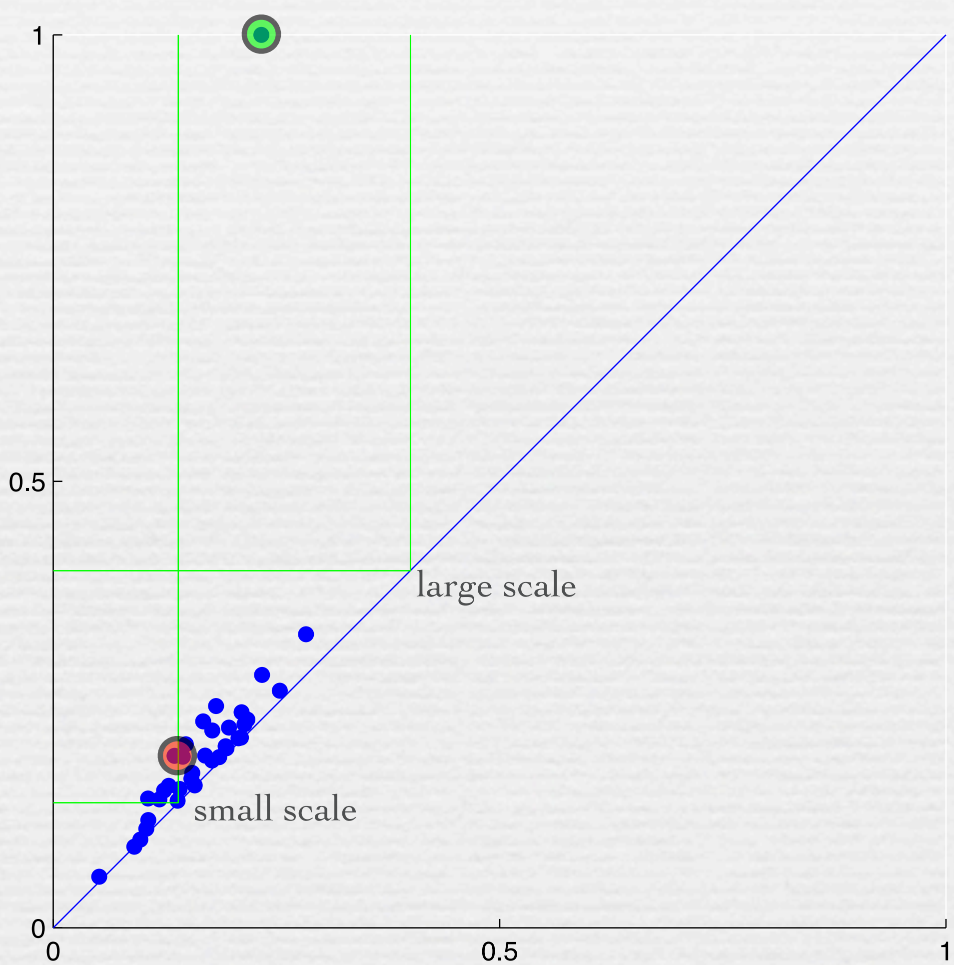
Persistence diagram

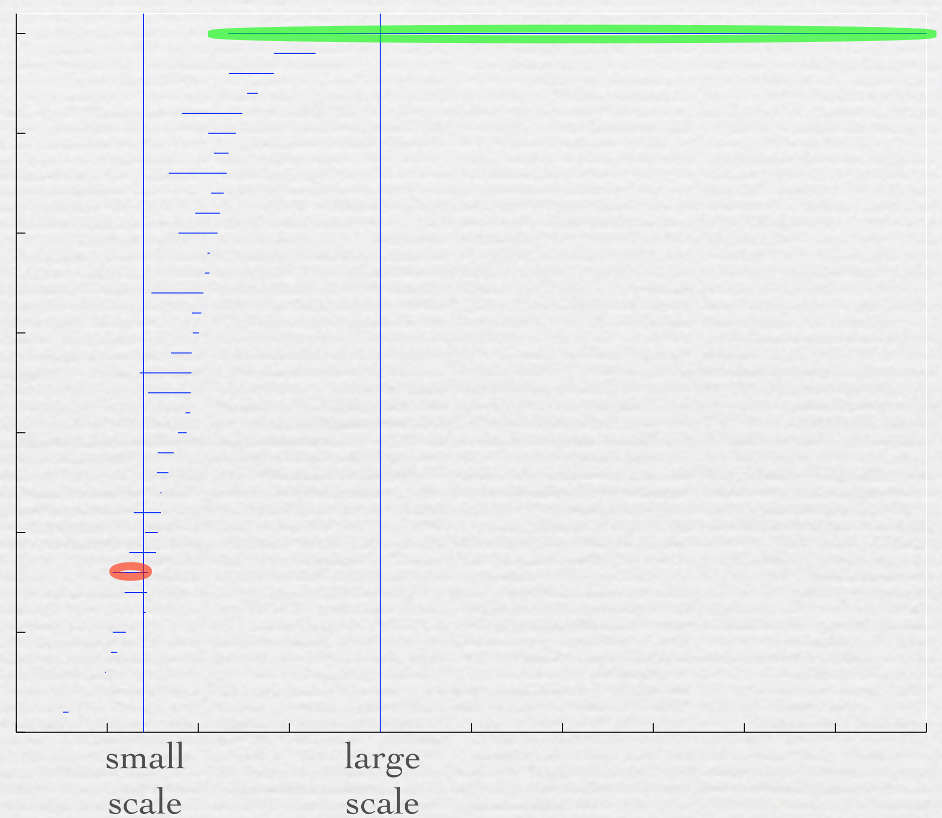
points (b,d)

Barcode

intervals [b,d)

Noisy circle





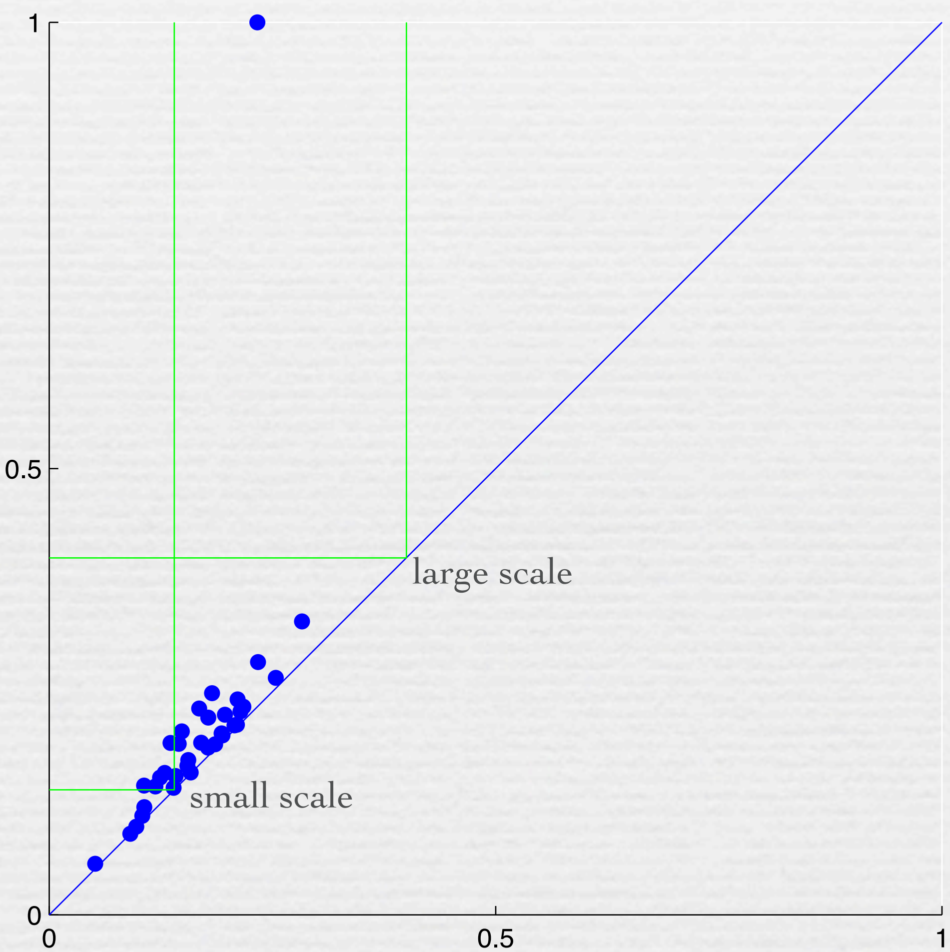
Persistence diagram

points (b,d)

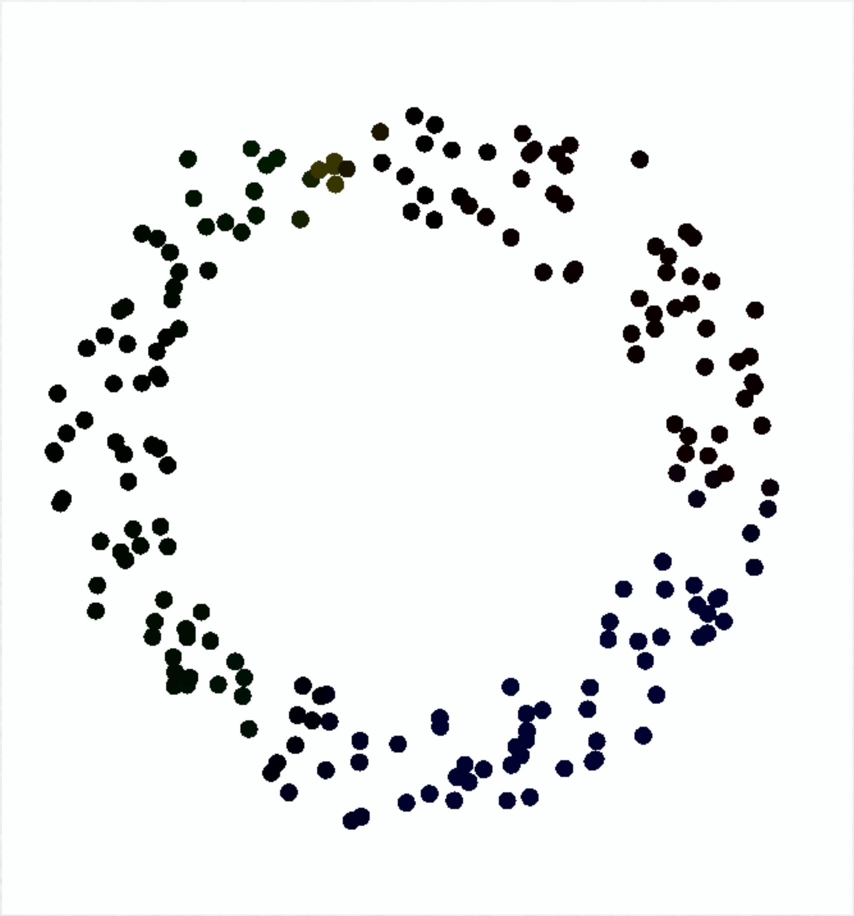
Barcode

intervals [b,d)

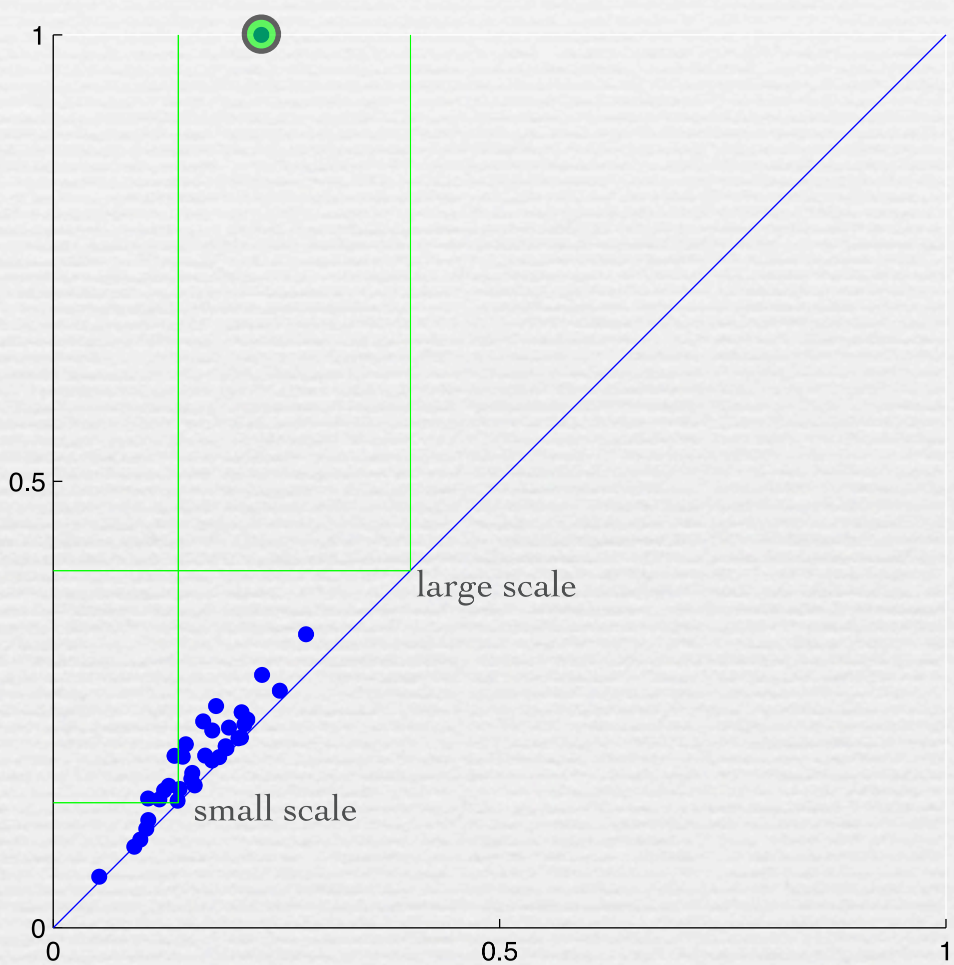
Noisy circle



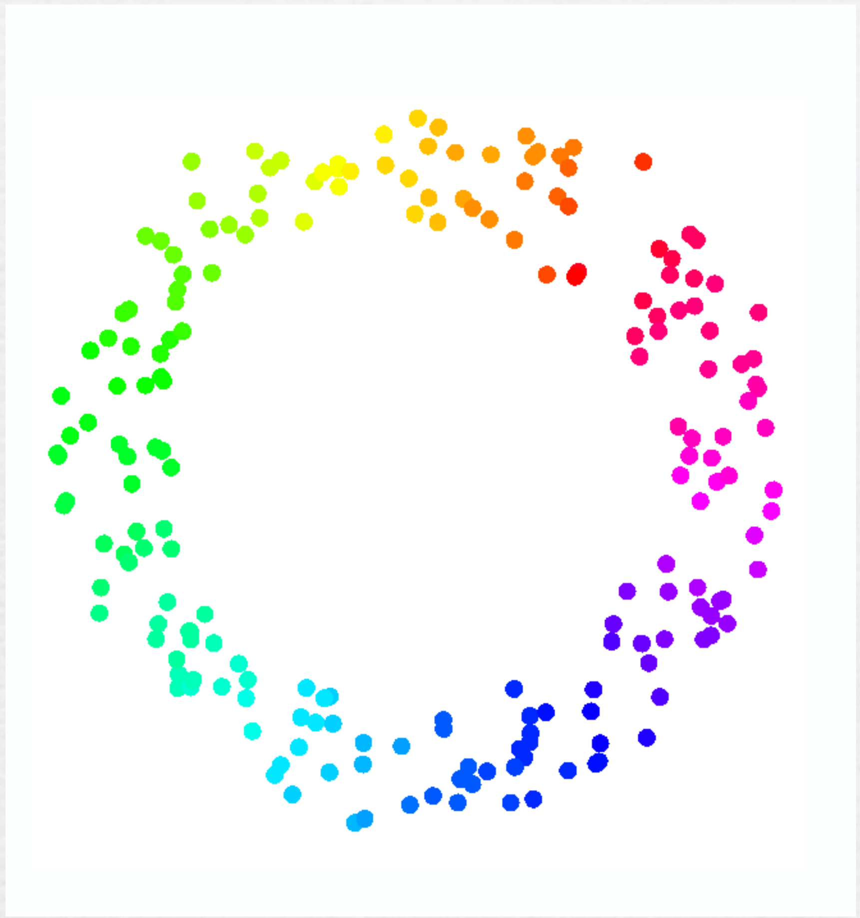
Persistence diagram



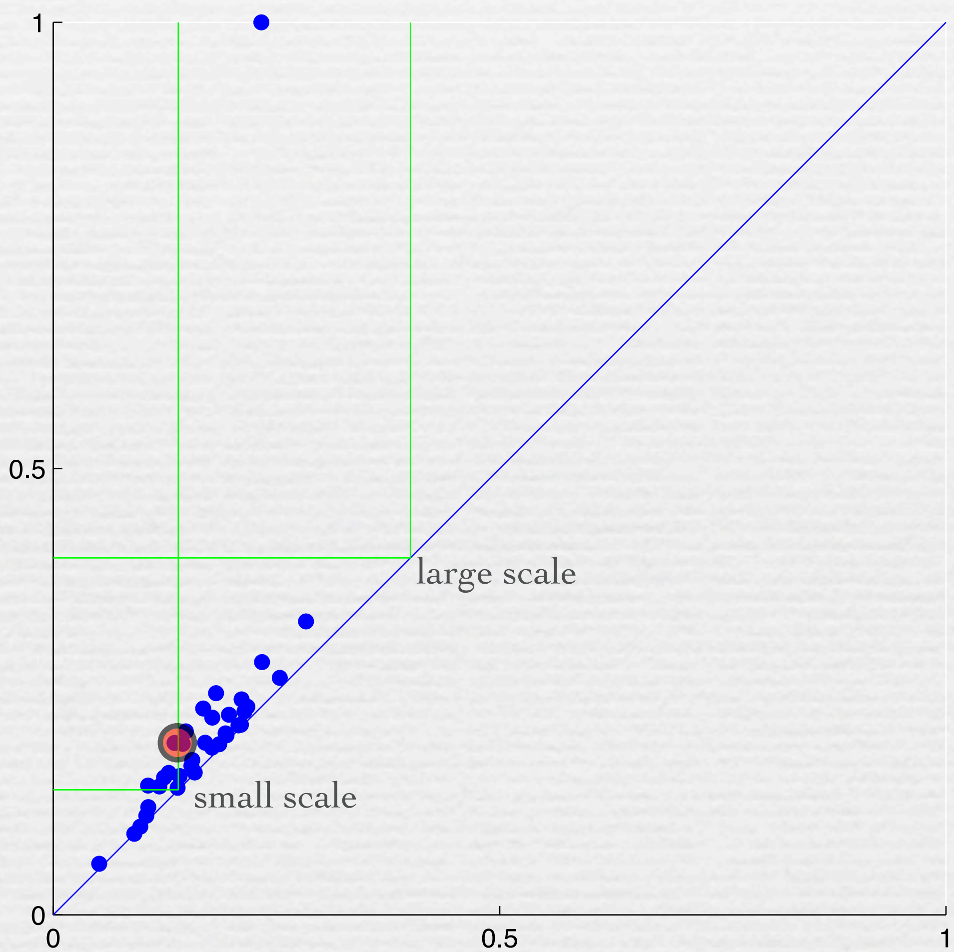
Noisy circle



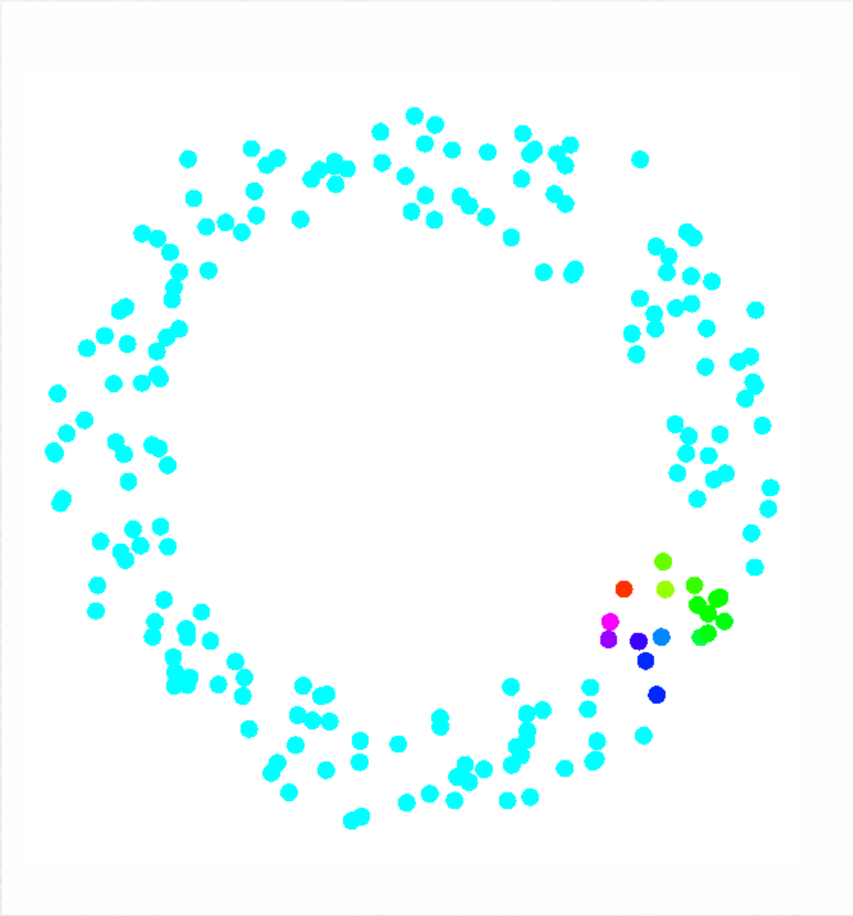
Persistence diagram



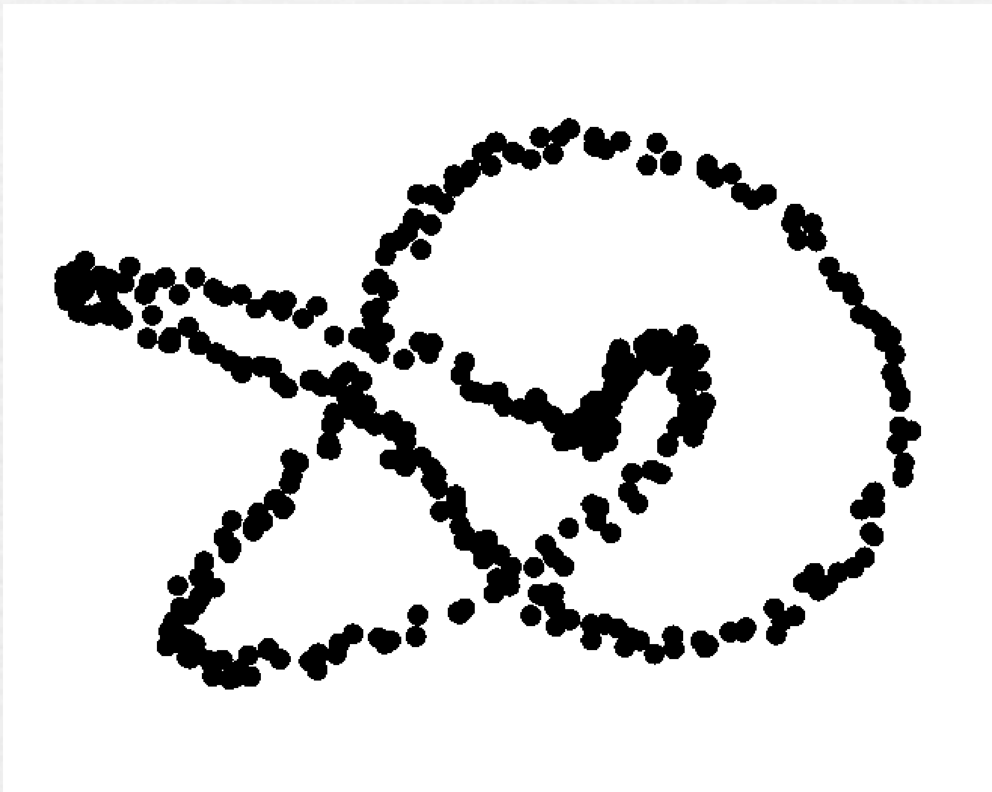
Noisy circle

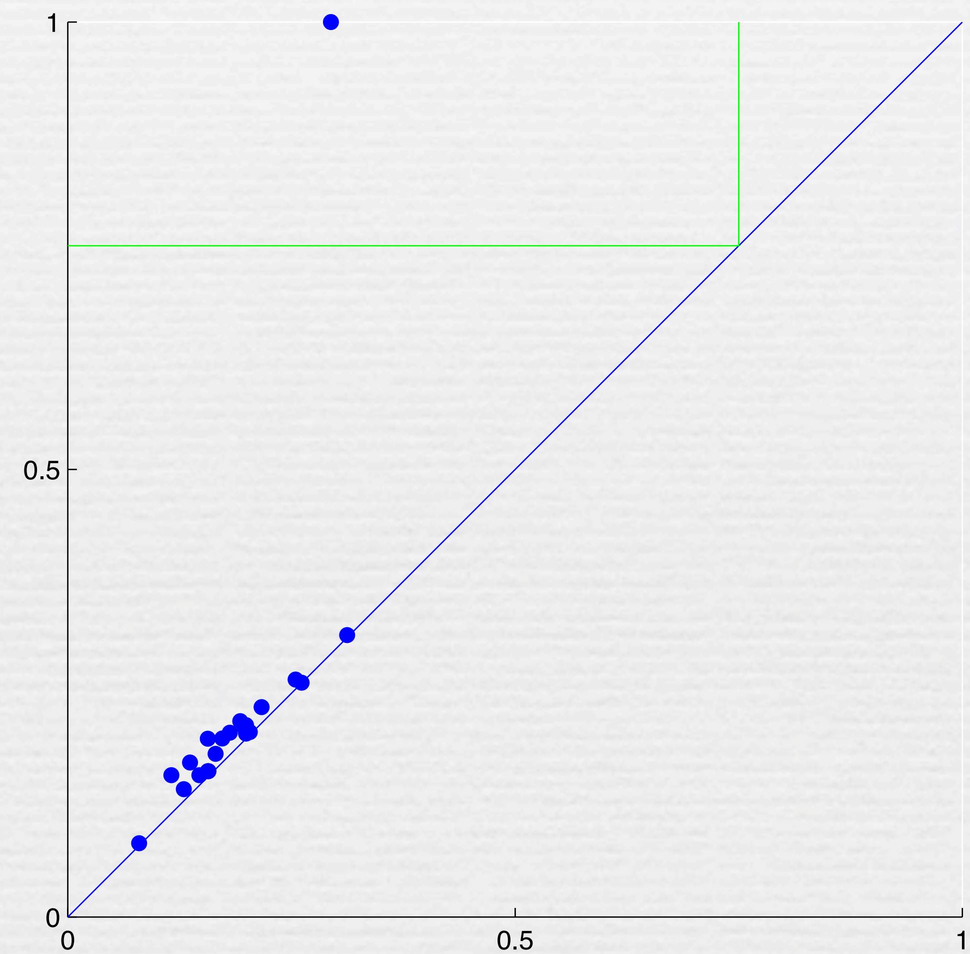


Persistence diagram

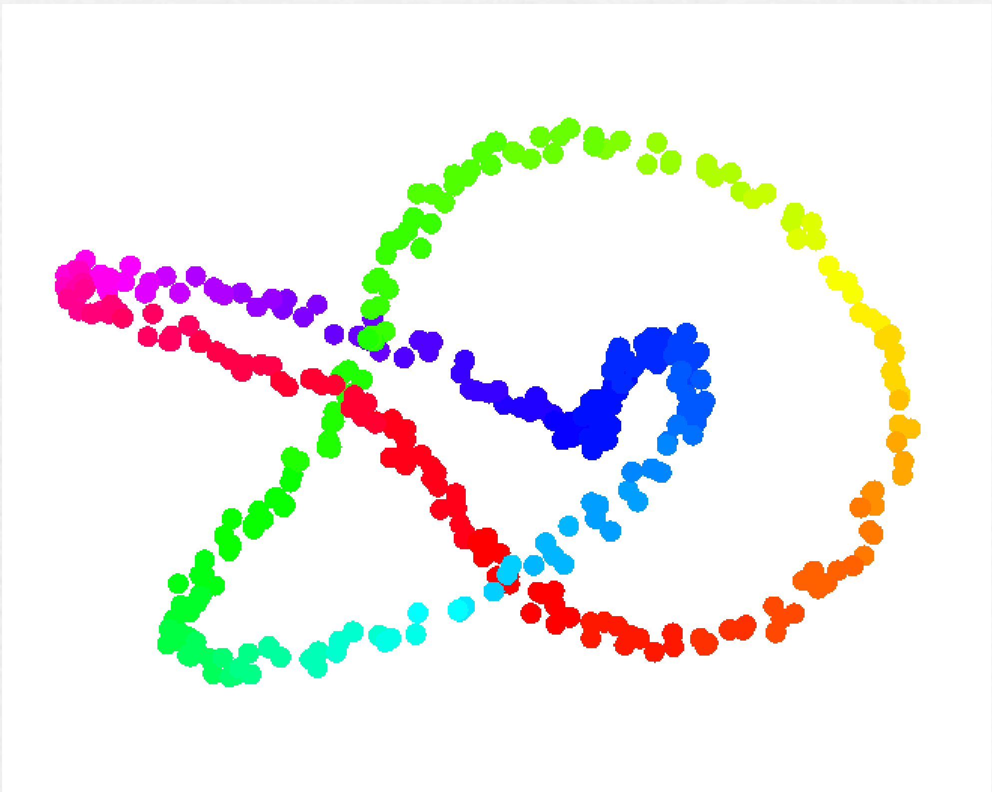


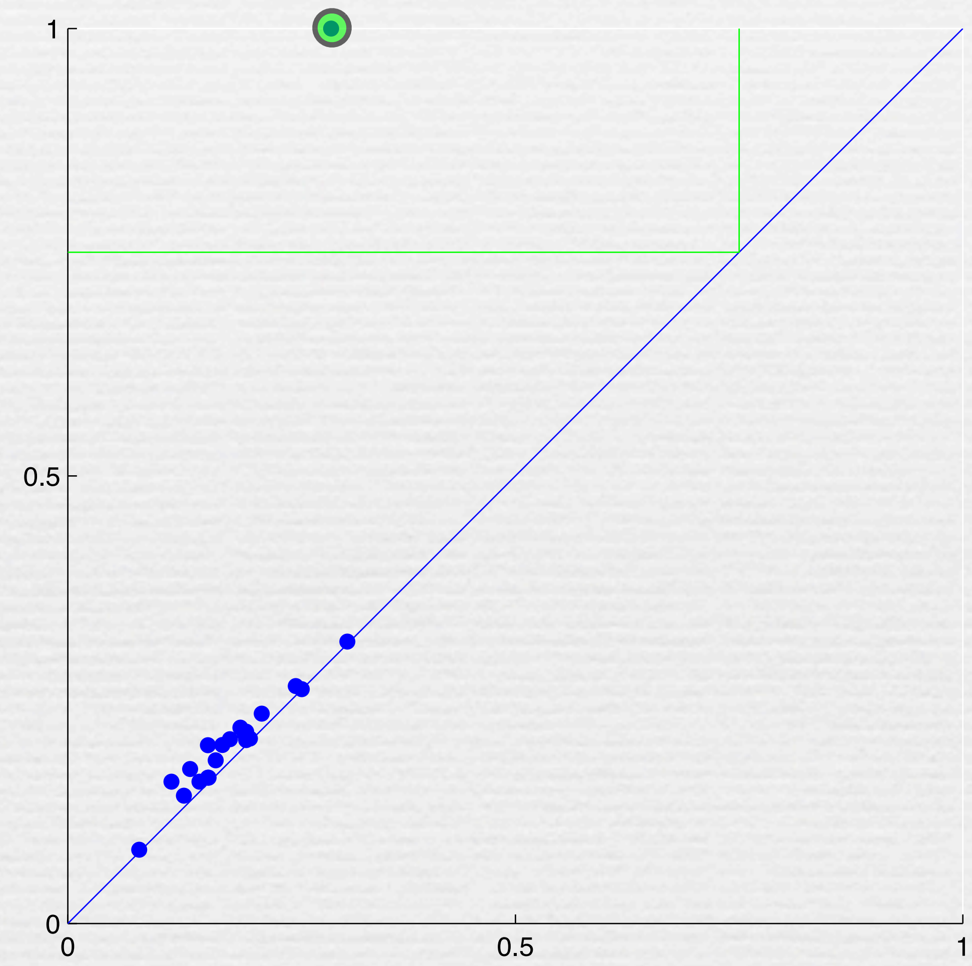
Trefoil knot



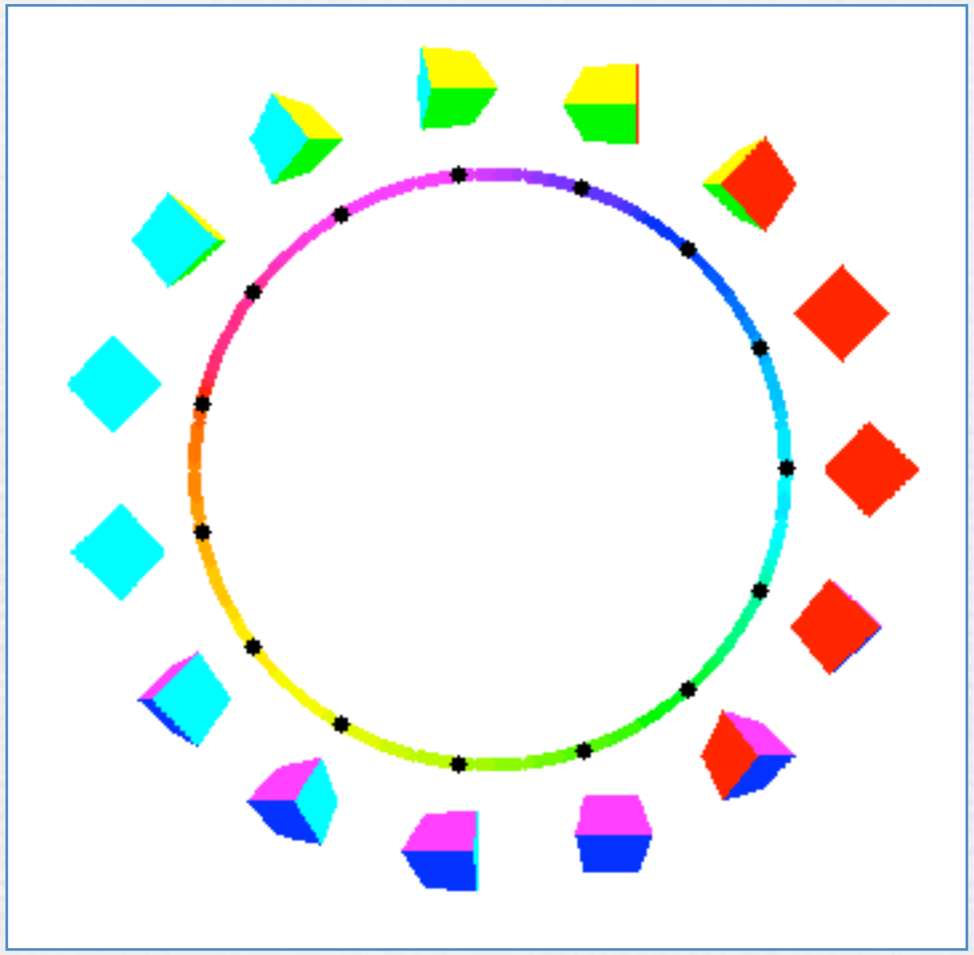


Trefoil knot

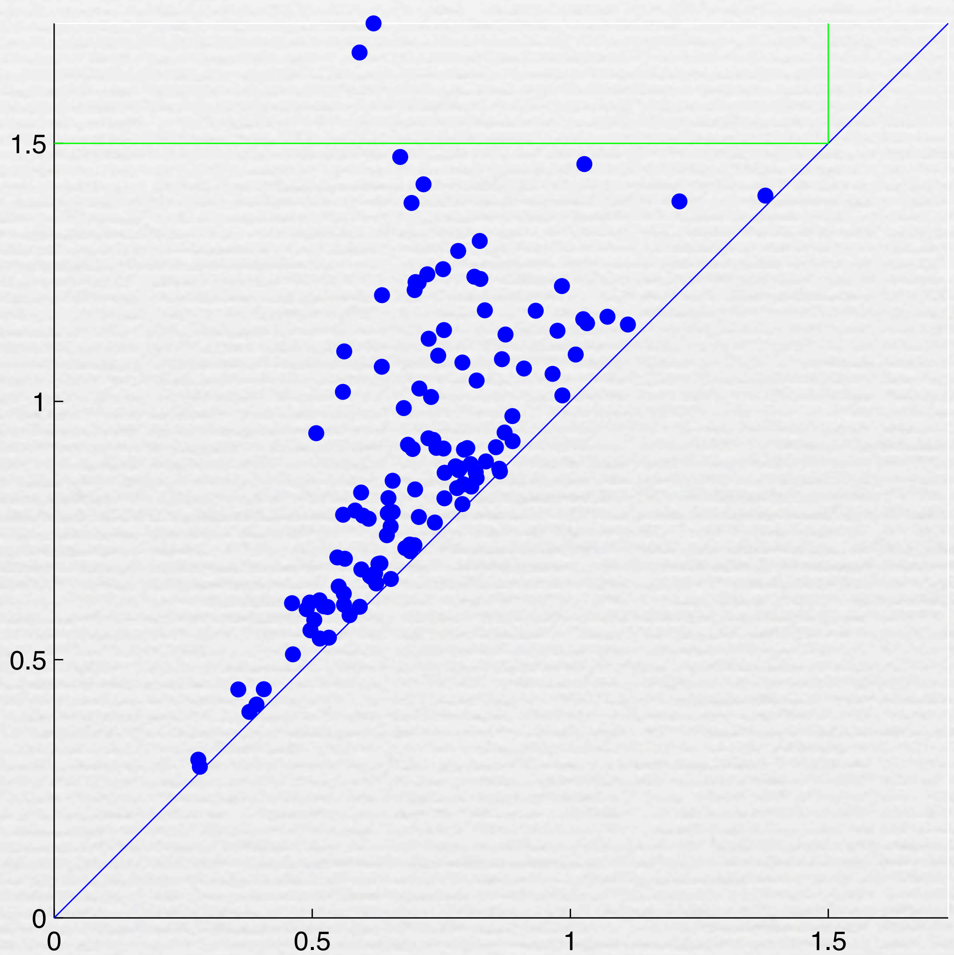


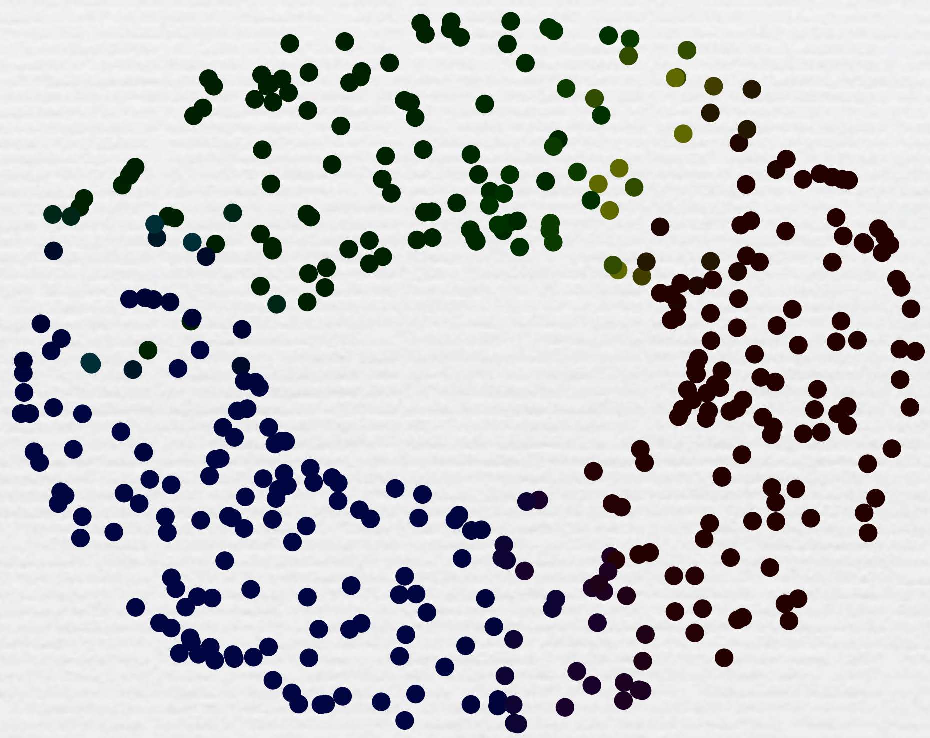


Rotating cube images

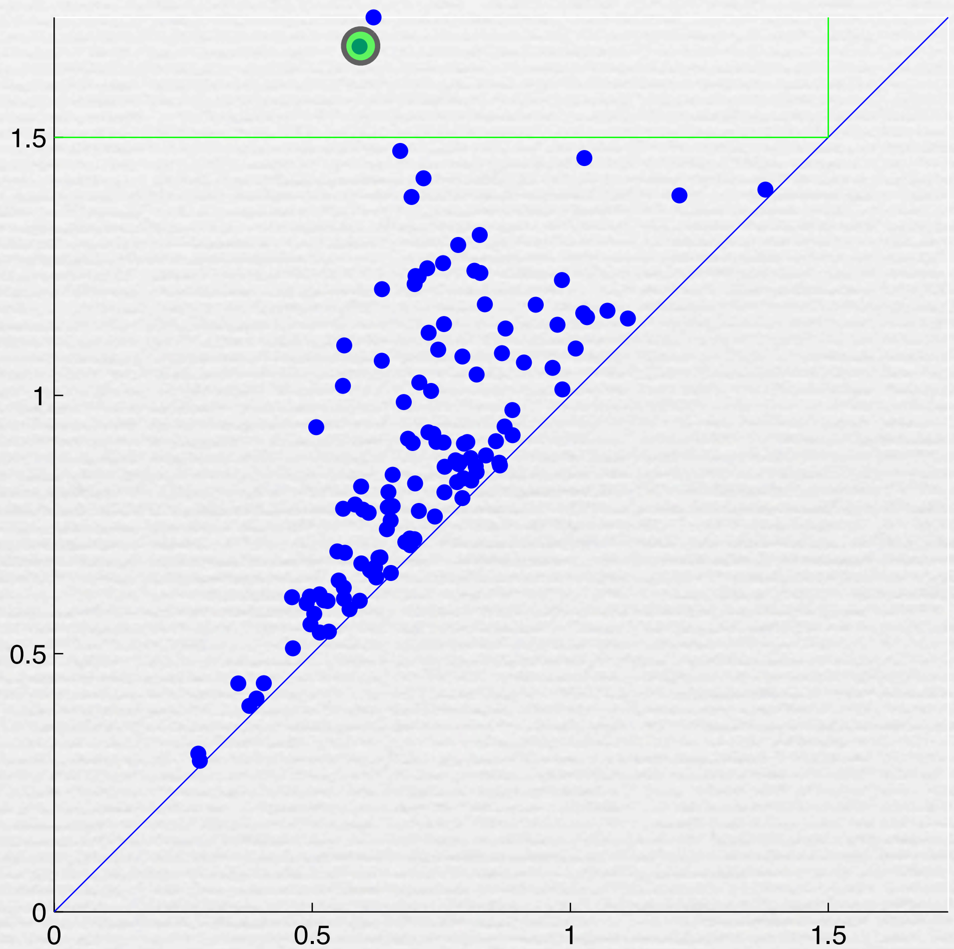


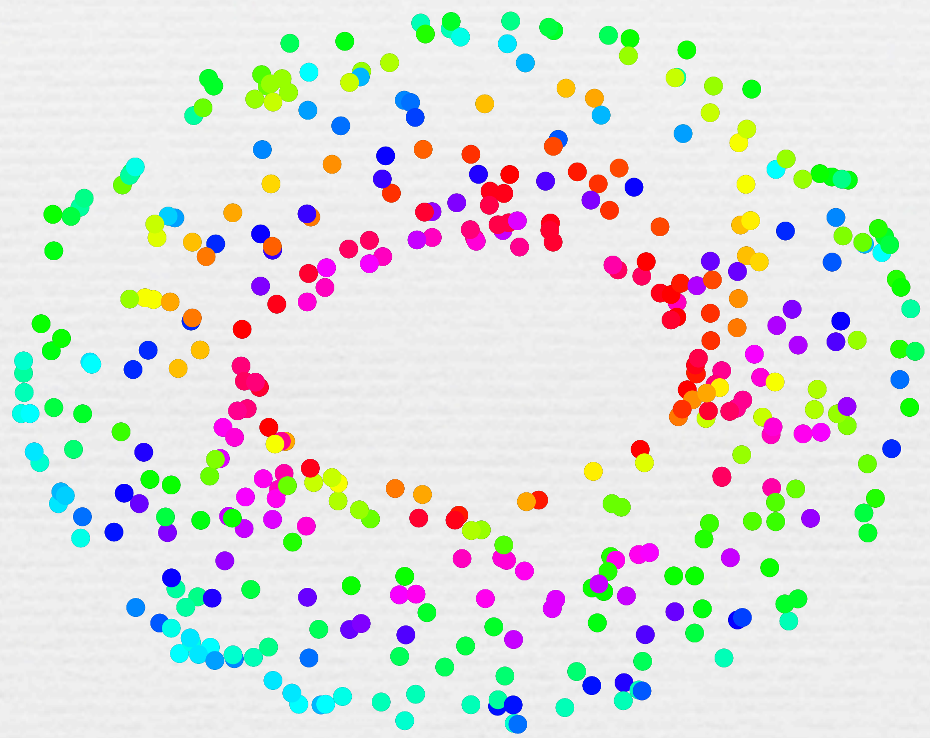
Torus



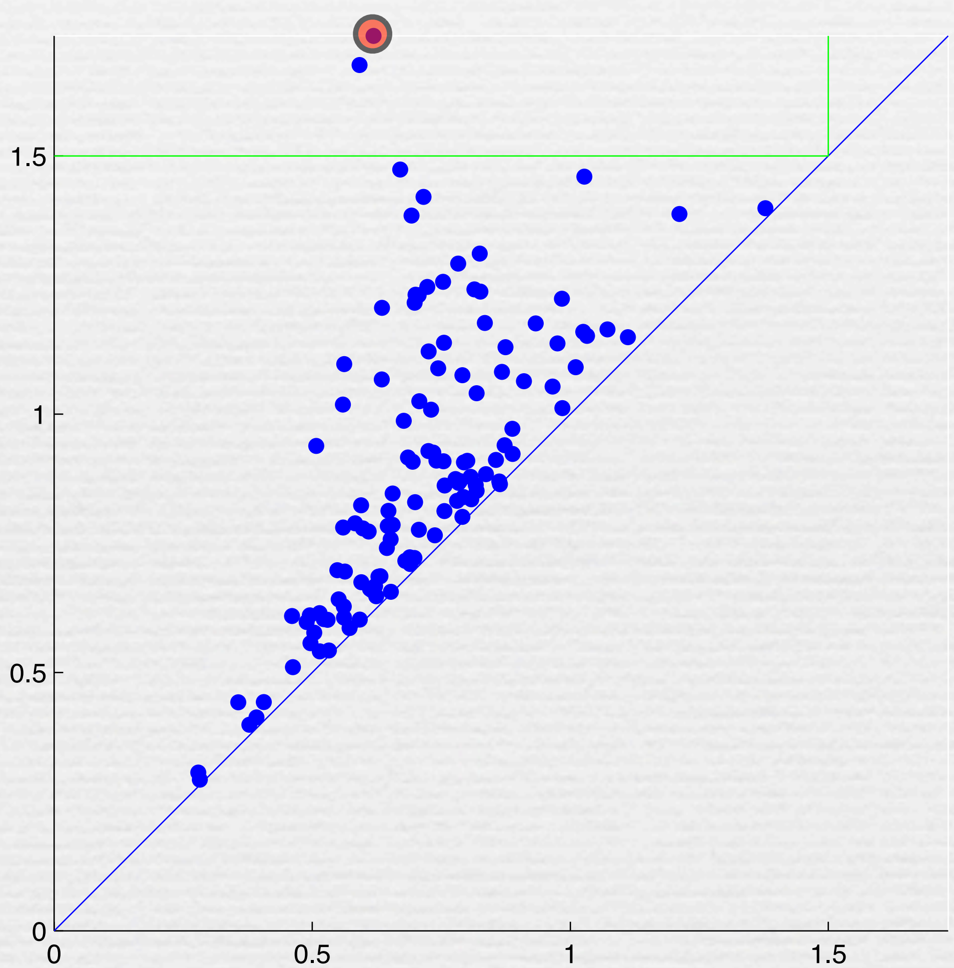


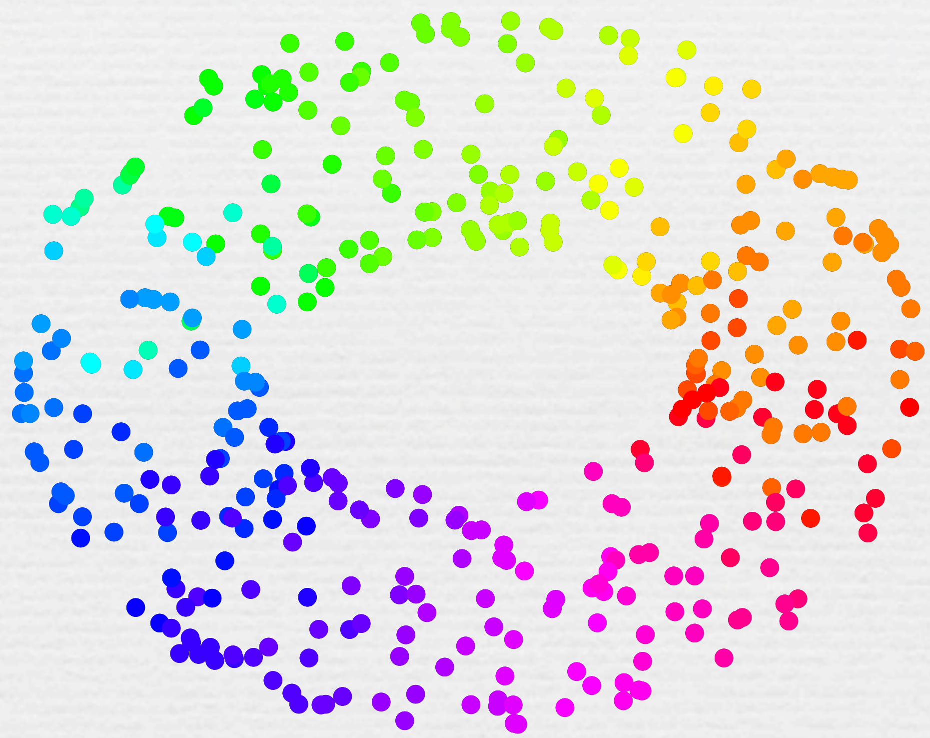
Torus





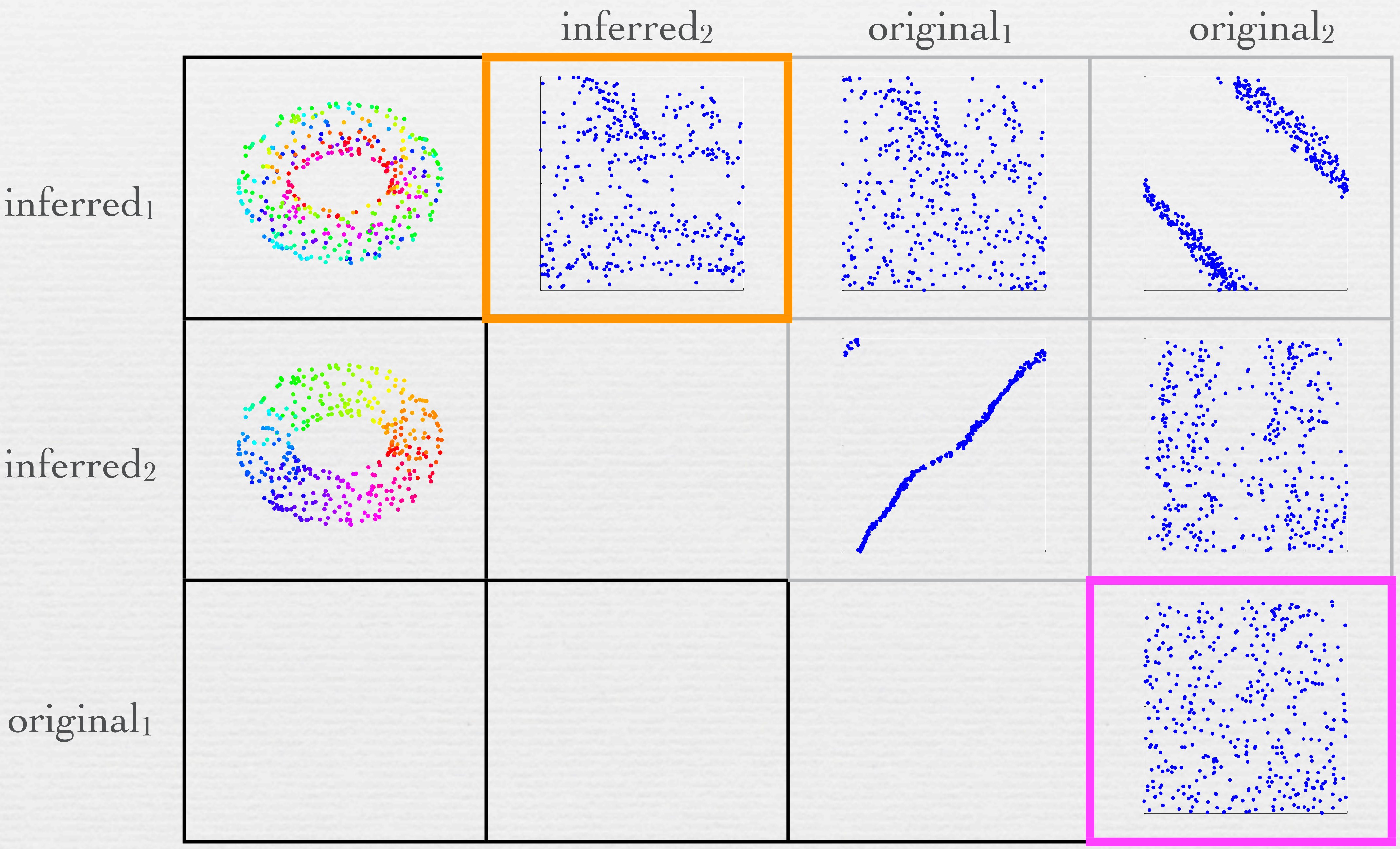
Torus





Wednesday, July 17, 13

Torus

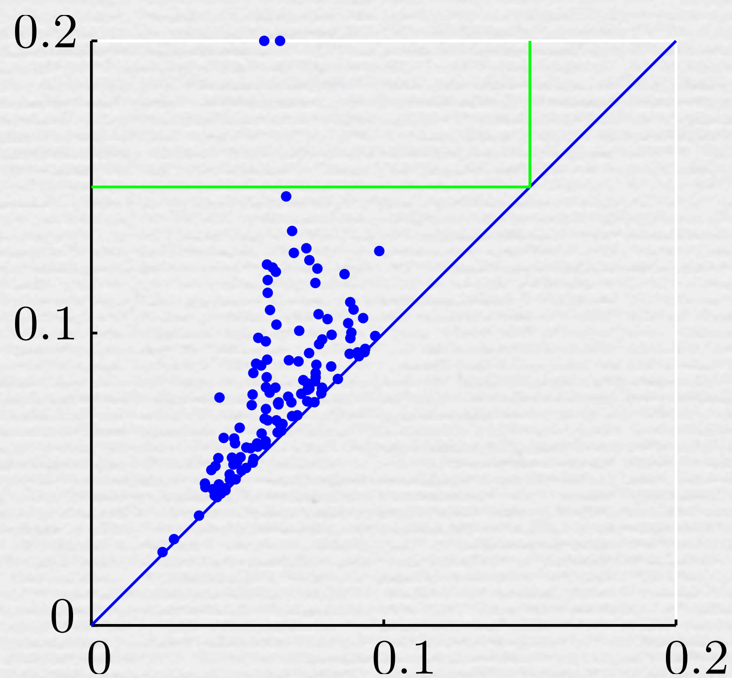

inferred$_2$
original$_1$
original$_2$
inferred$_1$
inferred$_2$
original$_1$

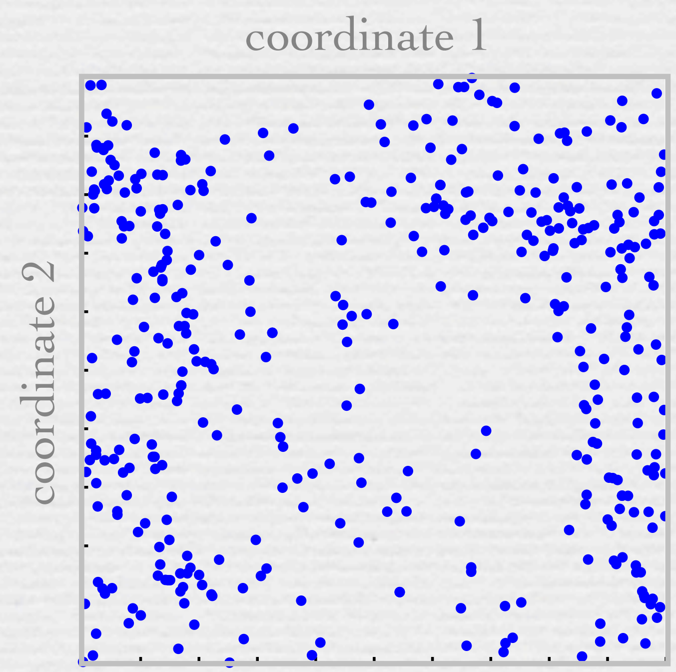
Elliptic Curve

- 400 points randomly chosen on

$$\{x^2y + y^2z + z^2x = 0\} \subset S^5 \subset \mathbb{C}^3$$

- Use projectively invariant metric $\mathrm{d}(\xi, \eta) = \cos^{-1}(|\xi \cdot \bar{\eta}|)$ to interpret as points in complex projective plane.

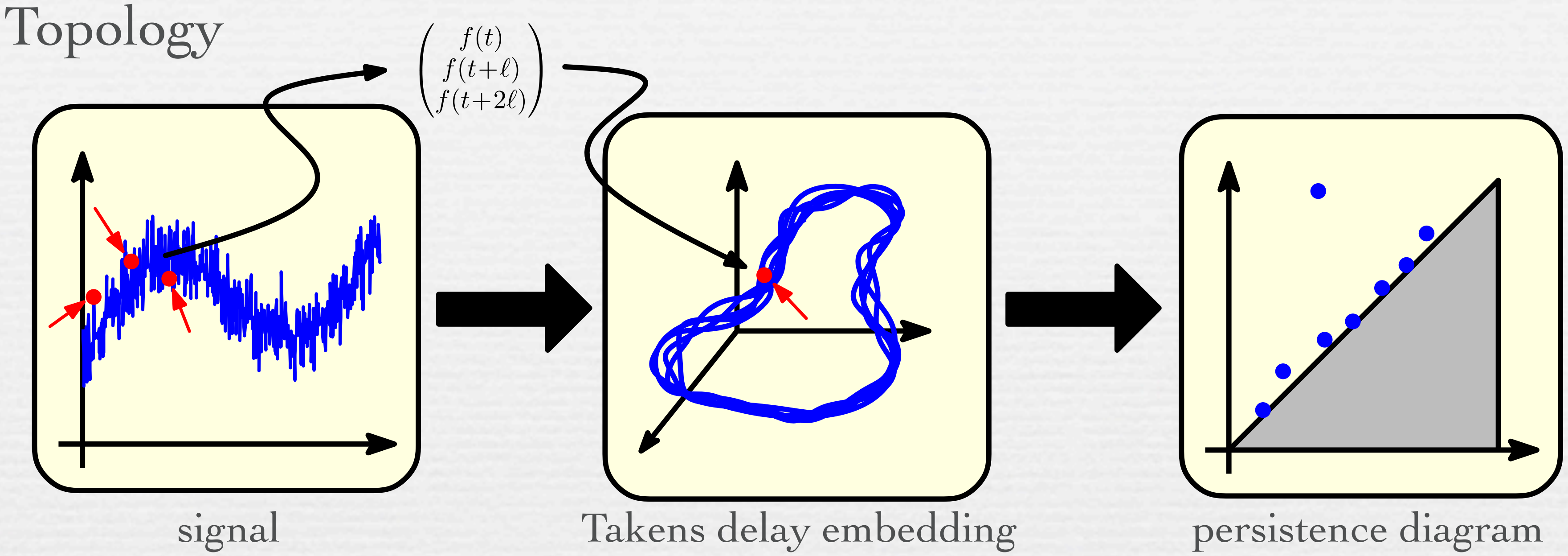




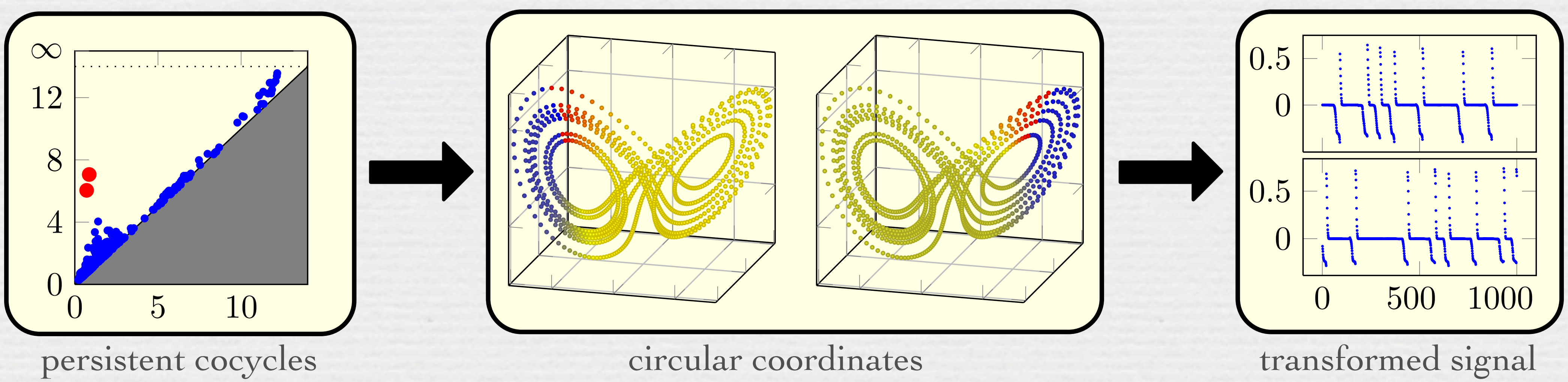
Time-series data

Topological analysis of time series

1. Signal to Topology

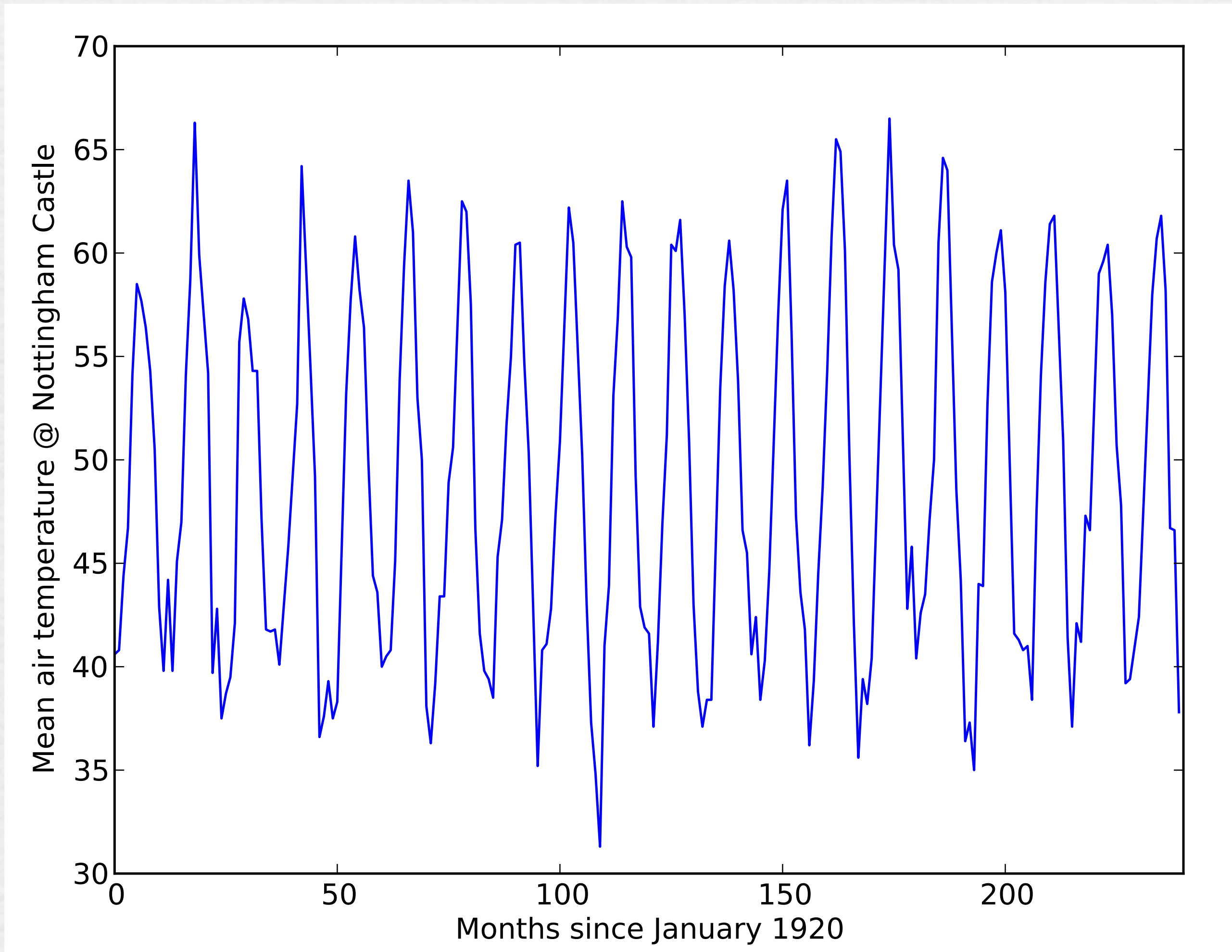


2. Topology to Transformed Signal



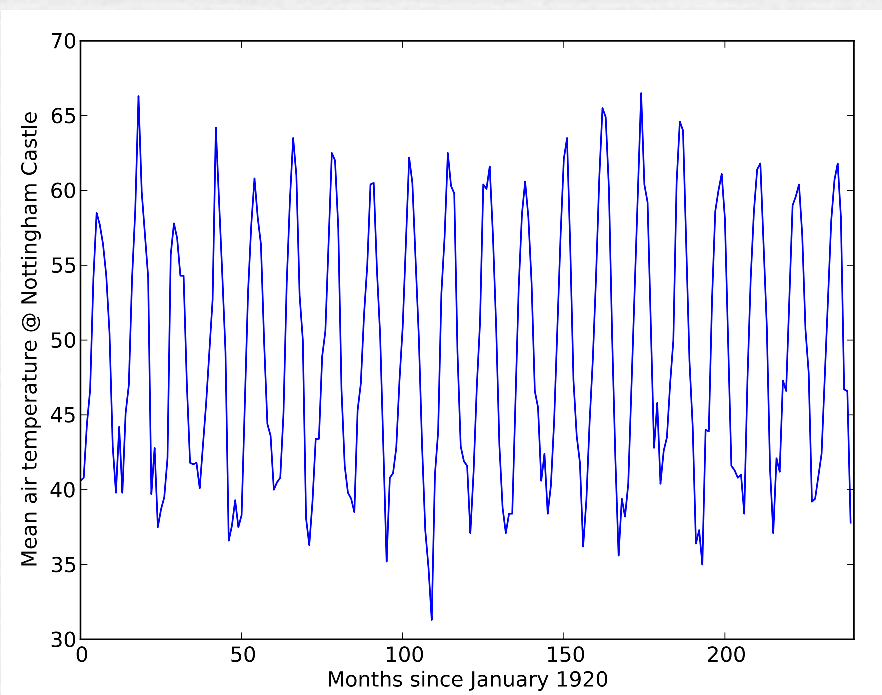
Period estimation

- Time series data:
  - Mean monthly air temperature at Nottingham Castle 1920-1939.
  - Source: O.D. Anderson (1976).

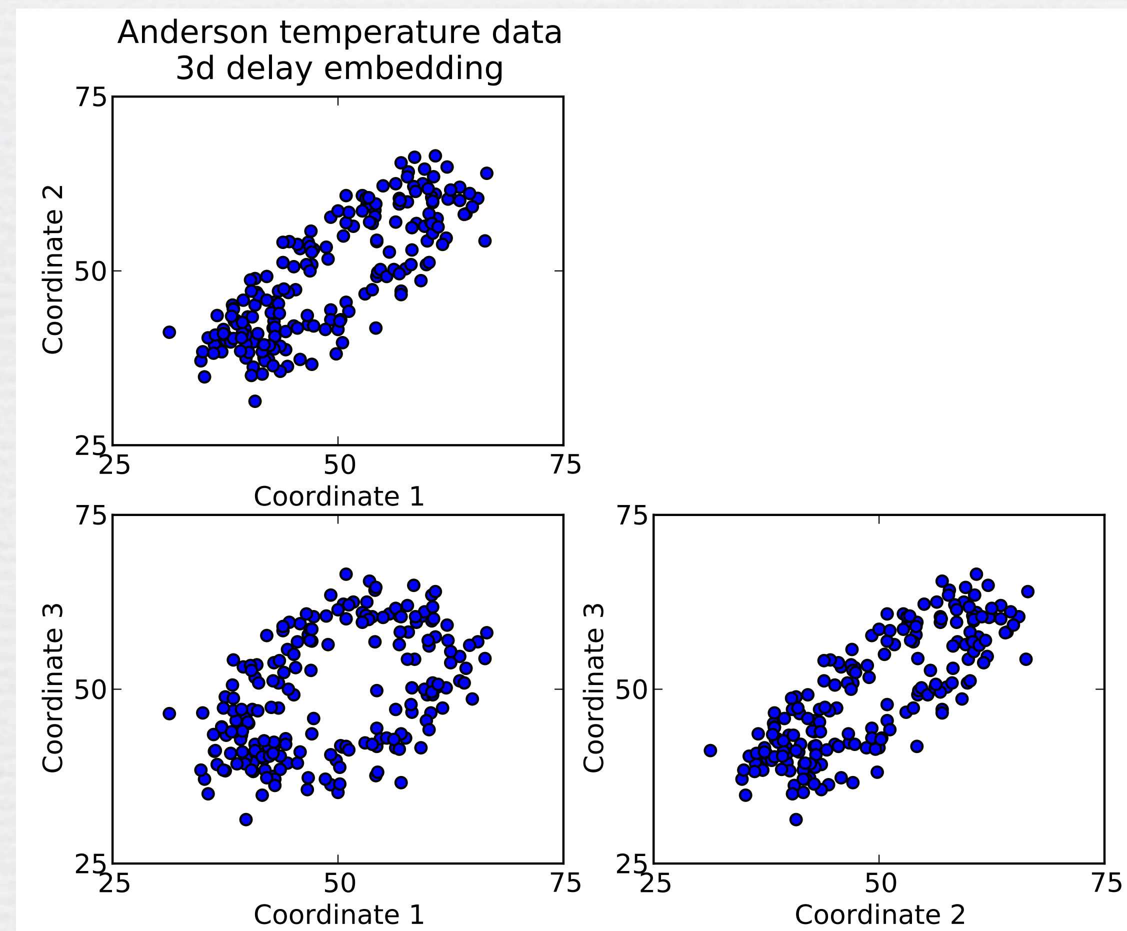


Period estimation

- Delay embedding (Takens 1981):
  - Convert 1-dimensional signal **a(t)** to n-dimensional signal **f(t)**.
  - **f(t)** = [**a(t)**, **a(t+k)**, **a(t+2k)**, ..., **a(t+(n-1)k)**].
  - Periodic signals remain periodic.
  - Circle topology may emerge in higher dimensions.

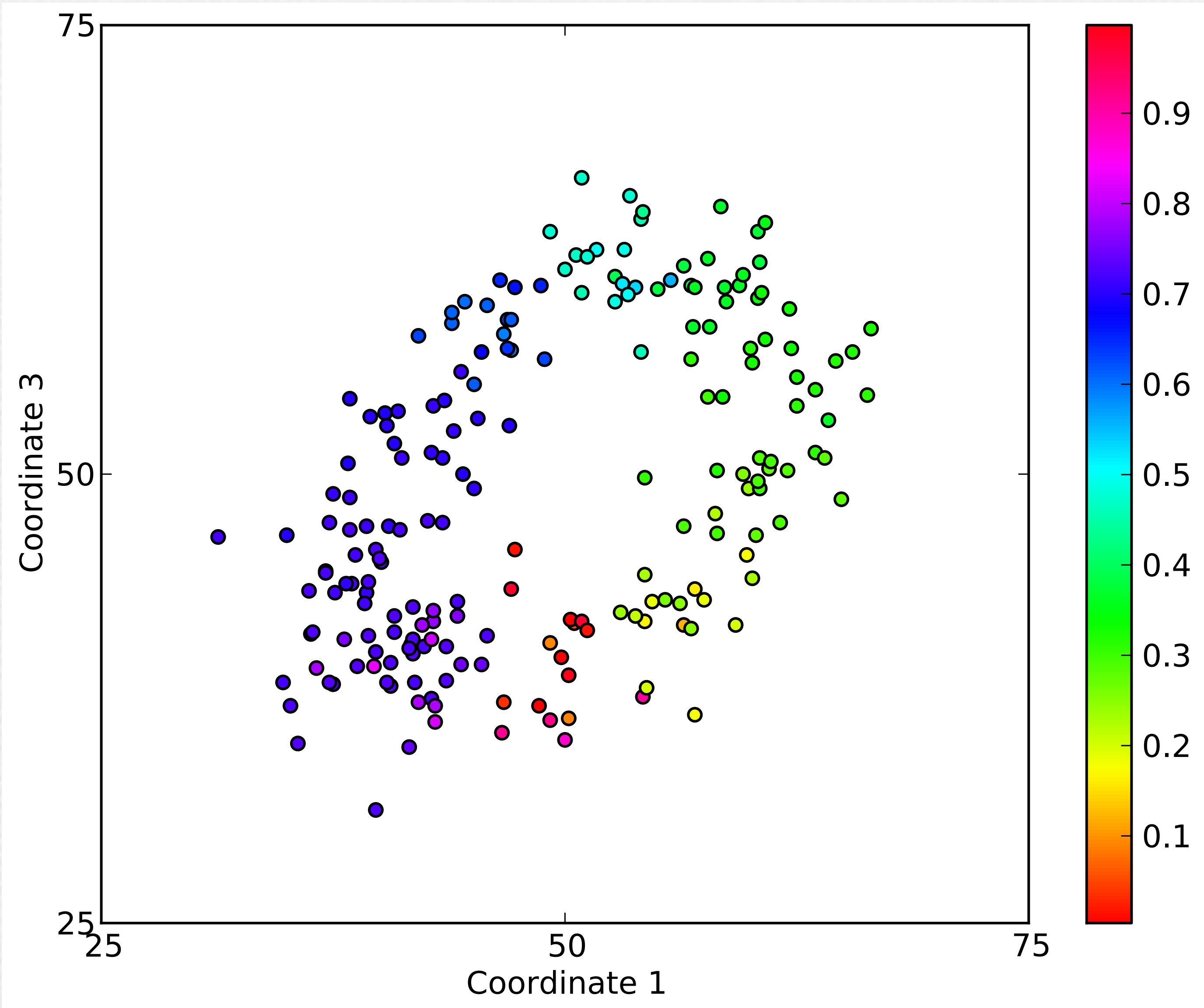


n=3

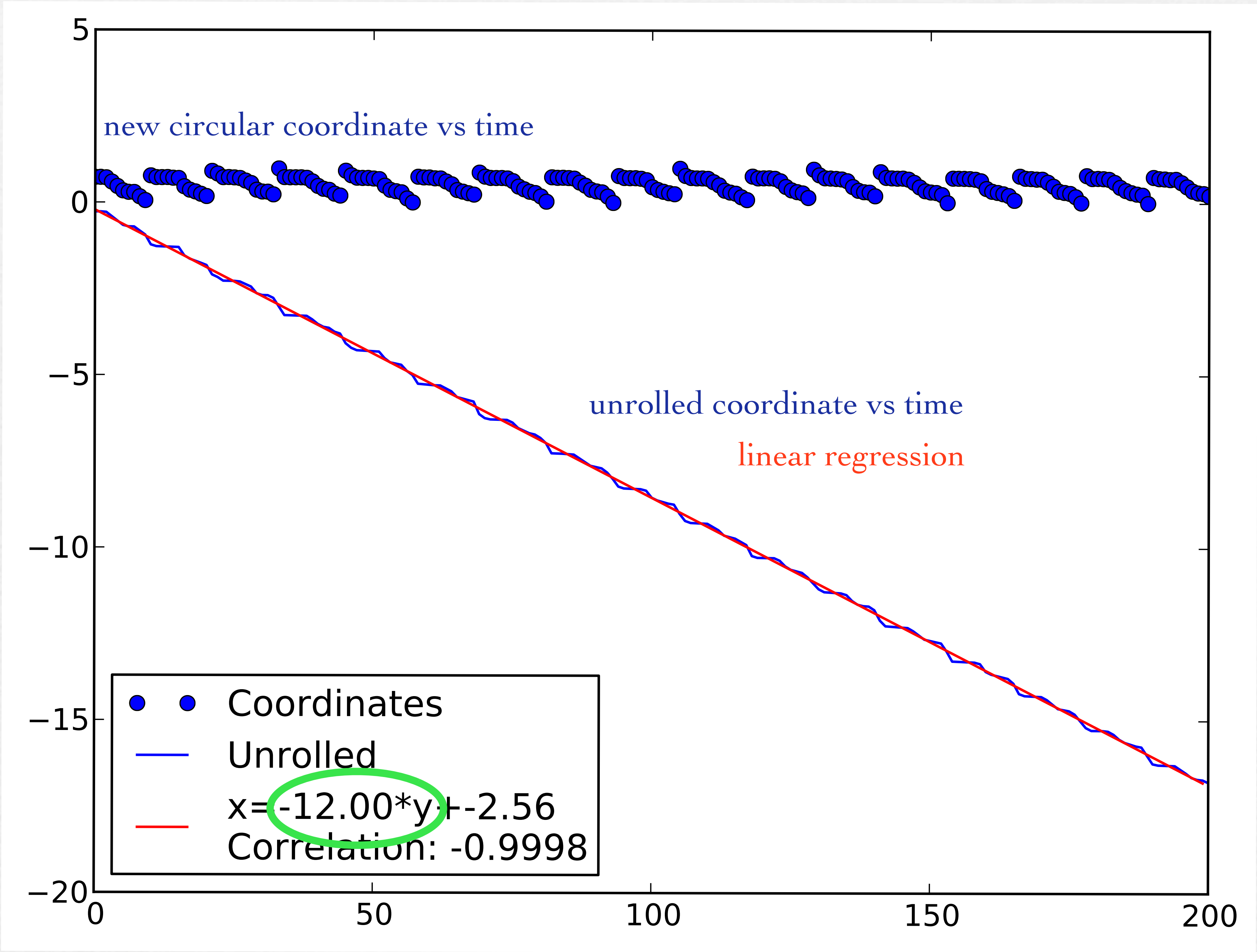


Period estimation

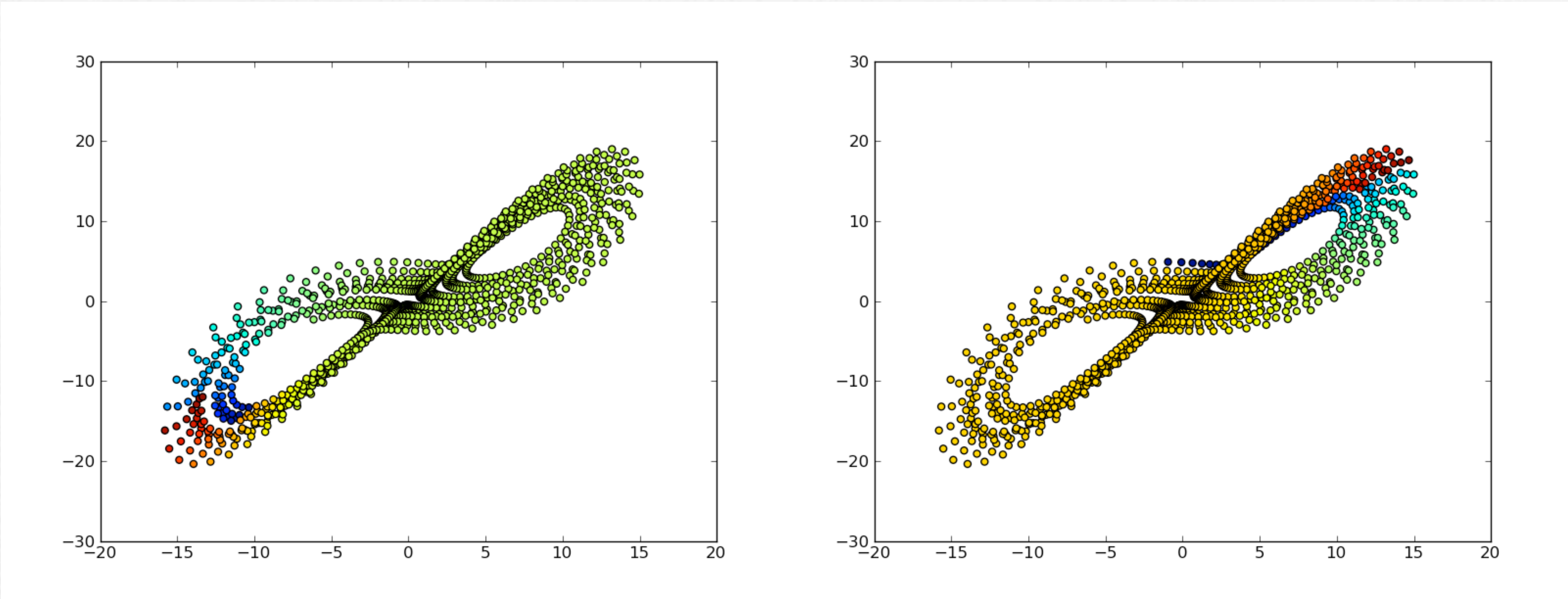
- Construct circular coordinate:



Period estimation

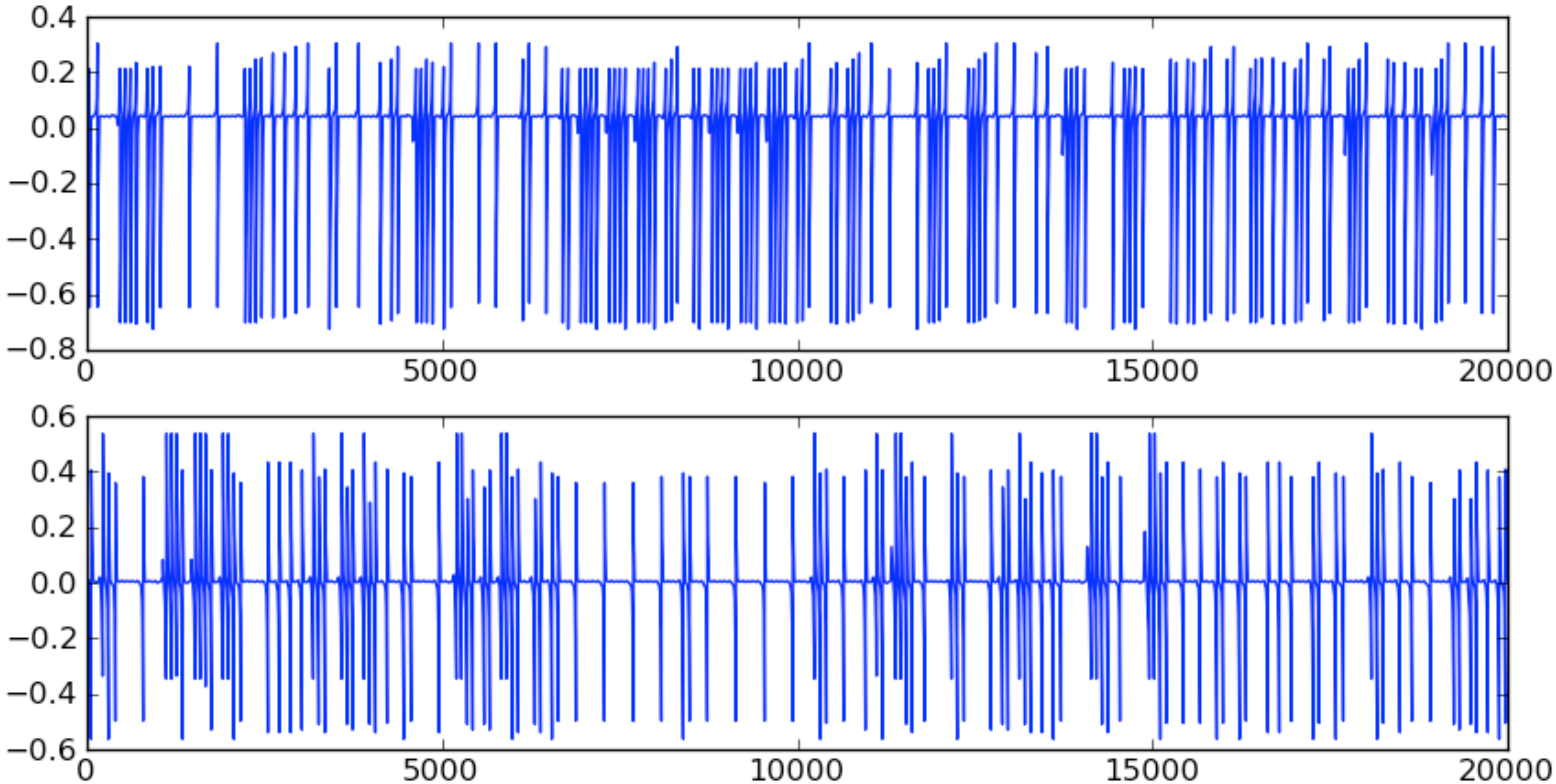


Lorenz attractor



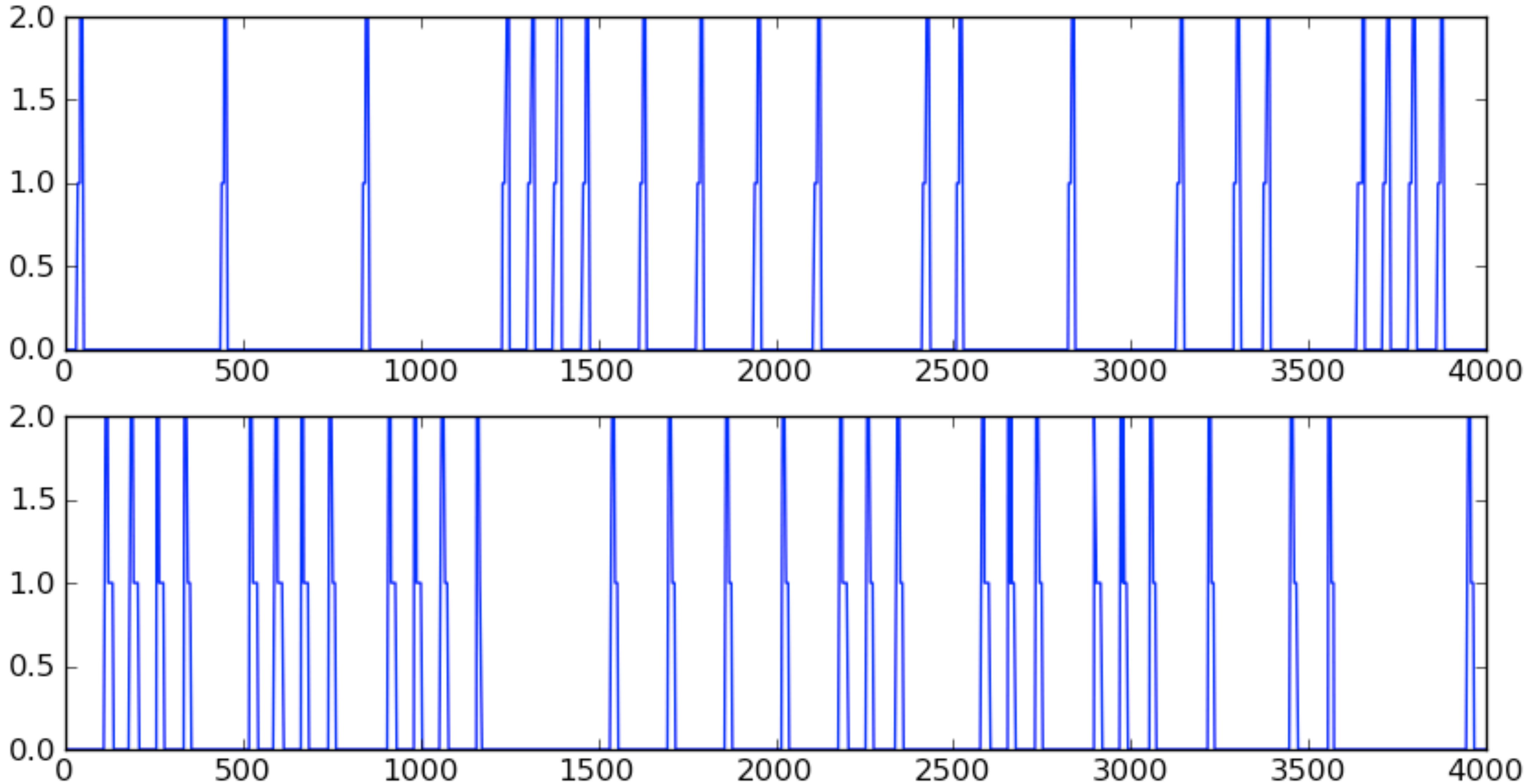
- Three-dimensional dynamical system.
- Data generated by following an arbitrary trajectory.
- Two cyclic coordinates found.
- We can track any other trajectory in terms of these coordinates.

Lorenz attractor



Lorenz attractor

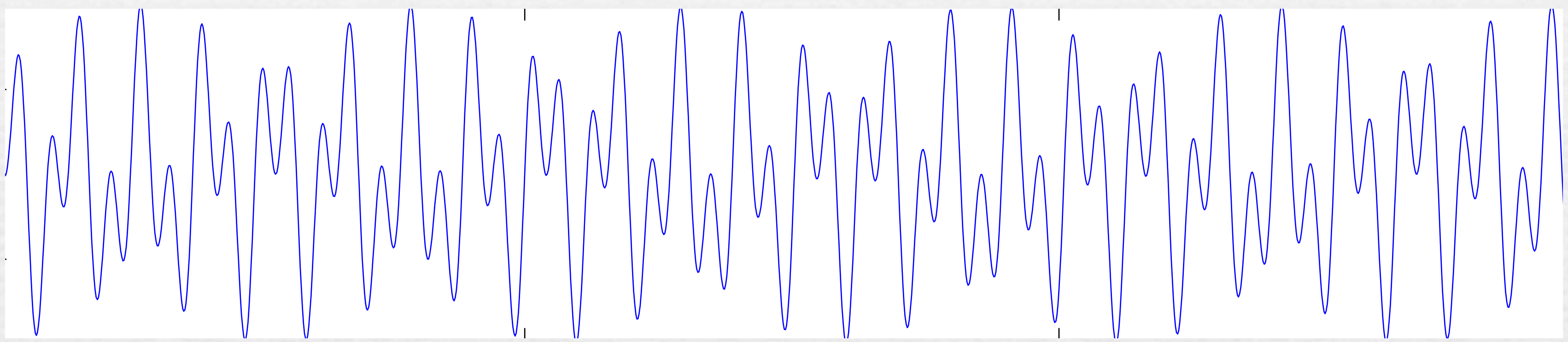
(Zoomed in and discretized)



Quasi-periodic signal

- Superposition of two signals:

$$f(t) = \sin(t) + \cos(\alpha t)$$



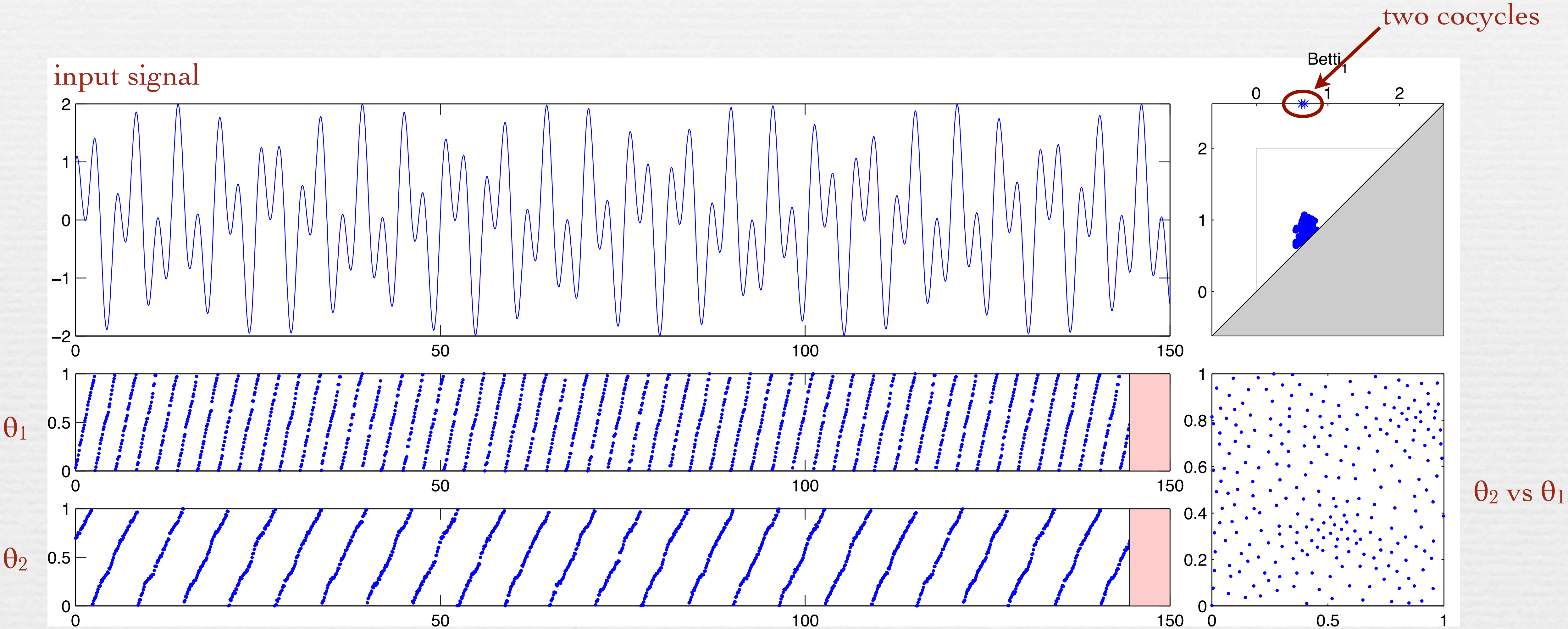
- If α is irrational, this converges to a dense sampling of a 2-parameter signal defined on a torus:

$$f(t) = \sin(\theta_1) + \cos(\theta_2) \qquad (\theta_1, \theta_2) \in (\mathbb{R}/2\pi\mathbb{Z}) \times (\mathbb{R}/2\pi\mathbb{Z})$$

- Takens embedding to recover the torus.

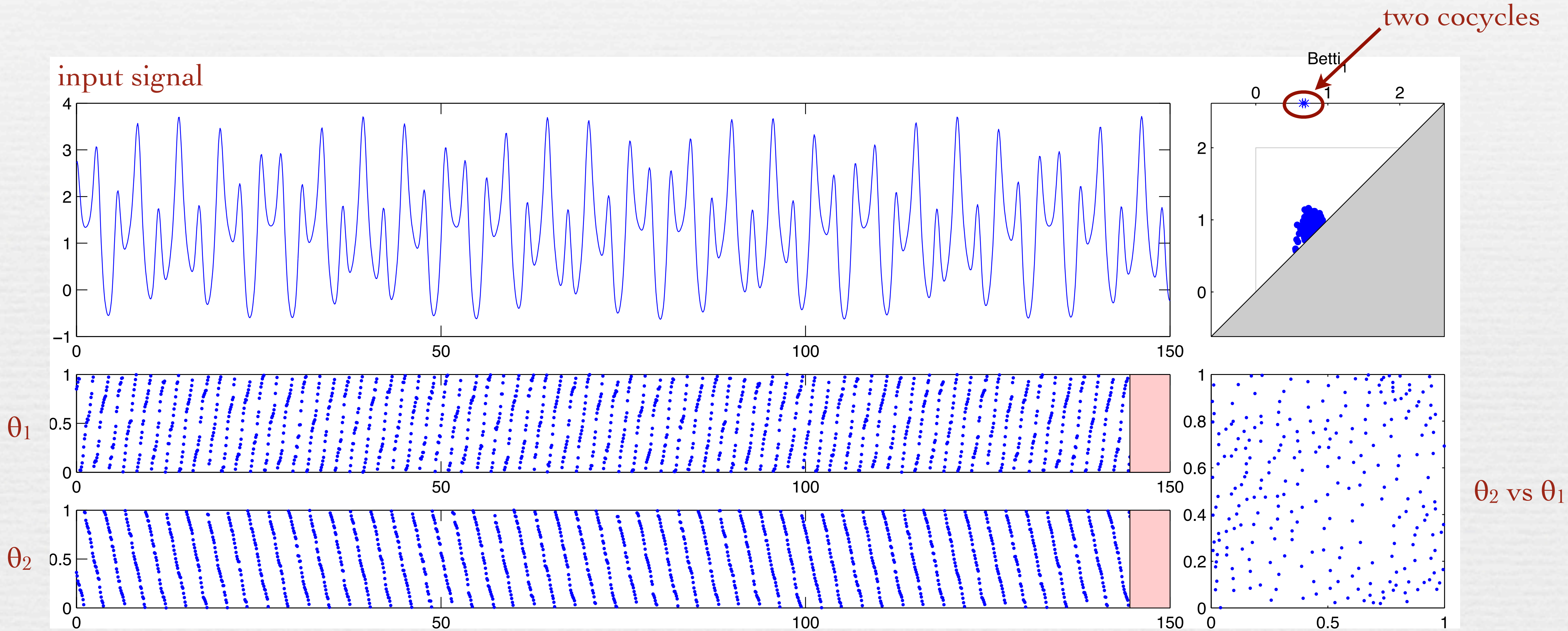
Quasi-periodic signal

$$f(t) = \sin(t) + \cos(\sqrt{5}t)$$



Quasi-periodic signal

$$f(t) = \sin(t) + \exp(\cos(\sqrt{5}t))$$



Wednesday, July 17, 13

Acknowledgements

- VdS partially supported by
  - DARPA HR0011-05-1-0007: TDA (Topological Data Analysis),
  - DARPA HR0011-07-1-0002: SToMP (Sensor Topology and Minimal Planning)
  - Digiteo Chair & project C3TTA
- MVJ partially supported by:
  - Office of Naval Research N00014-08-1-0931
- Thanks to:
  - Dmitriy Morozov, Konstantin Mischaikow, Yuliy Baryshnikov
  - Joshua Tenenbaum, Gunnar Carlsson, Robert Ghrist, Frédéric Chazal
  - Pomona College, Stanford University, INRIA Saclay – Île-de-France