49

Determining Periodicity in Data

Kelly Spendlove
Department of Mathematics, Rutgers University

background

Description

- High-throughput technologies can collect massive amounts of data often being generated by an unknown nonlinear system
- Both the *geometry* of the data and the *action* of the unknown nonlinear system are of interest

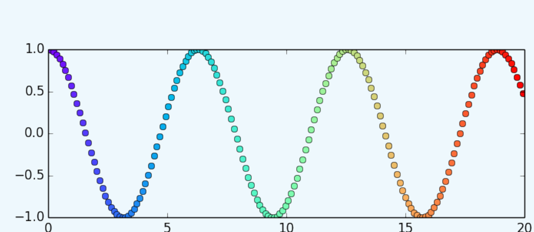
Challenges

- Nonlinear systems can exhibit complex behavior *at all scales* with respect to parameter and phase space
- In contrast: experimental data is typically noisy, of limited precision and the underlying system may be chaotic

Thus when attempting to understand nonlinear phenomena we must *extract relatively crude invariants of the data to compare to any model*

problem statement

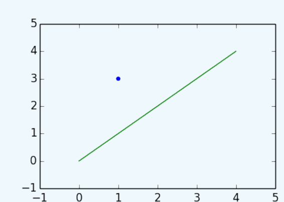
- Given these challenges we start in one of the simplest situations: determining whether the data is periodic
- Formally, a function f is periodic if there exists a period P such that $f(x+P) = f(x)$ for all P



the shape of data

Persistent homology extracts multiscale structure of data by progressively fattening the points in a point cloud

Once the balls are a certain radius a hole is *born* in the example above. At a large enough radius the hole *dies*. This is encoded in the *persistence diagram*.



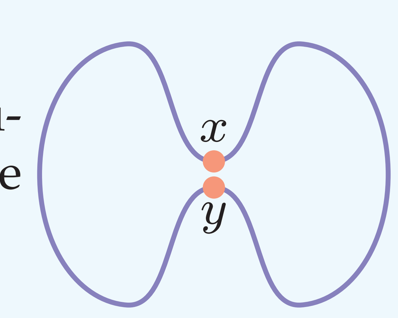
The persistence diagram represents the birth and death of a hole by the ordered pair (b, d)

amplifying topological signals

The diffusion distance provides a different means of calculating the distances (and persistence diagram) between points in the point cloud

- Intuitively two points are close in the diffusion distance if there are many short paths to get from one point to the other
- Two points are far in the diffusion distance if there are few paths to get from one point to the other

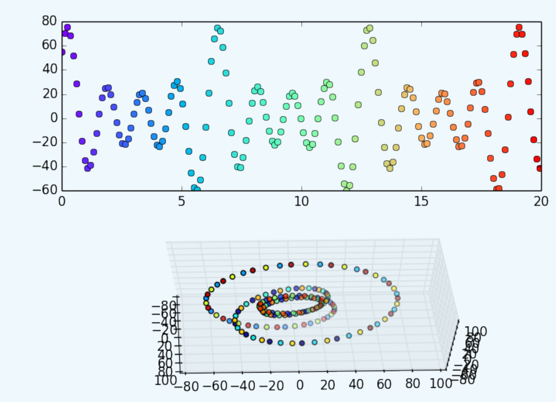
x and y are close in the Euclidean distance but far in the diffusion distance



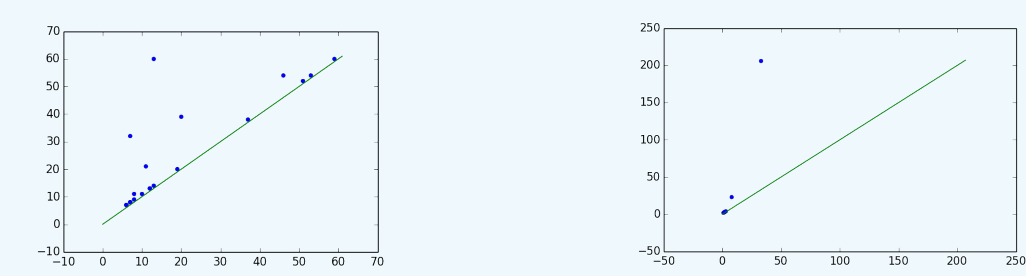
A persistence diagram calculated with diffusion distance is a *diffusion diagram*.

diffusion distance

- Trigonometric polynomial embedded in three dimensions



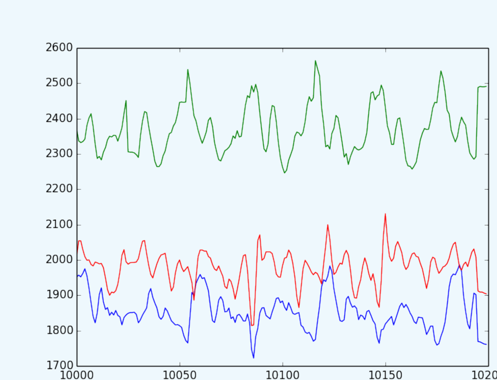
- Diffusion amplifies topological signal over Euclidean distance



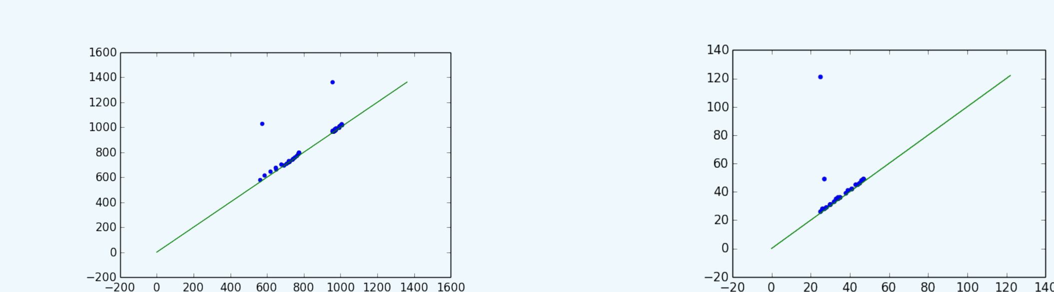
(left) persistence diagram (right) diffusion diagram

high embedding dimensions

- A high embedding dimension seems to smooth the data
- Accelerometer data from UCI data repository [2]
- Diagrams for y (green) time series using embedding dimension 100

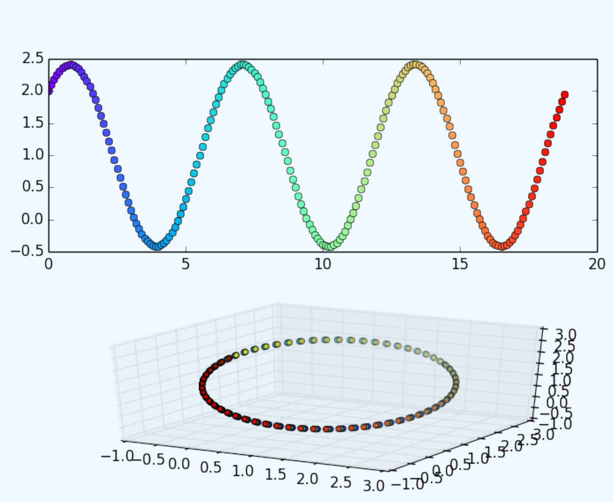


time series of acceleration while walking (x, y, z coordinates)

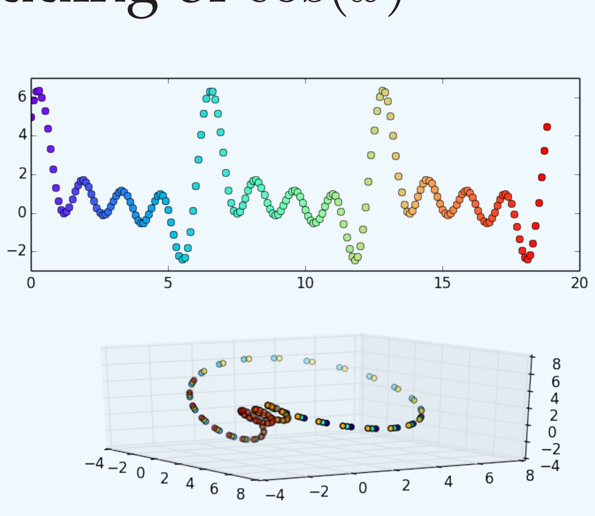


(left) persistence diagram (right) diffusion diagram

takens' embedding theorem



embedding of $\cos(x)$



embedding of trigonometric polynomial

- Takens' embedding is a transformation of a time series $(x_1, \ldots, x_n)$ to a point cloud
- For fixed integer lag τ and embedding dimension m the map is given by

$$\Phi(x_i) = (x_i, x_{i+\tau} \ldots, x_{i+\tau m})$$

- For a *generic* choice of these parameters the point cloud will be embedded as a (topological) circle if the time series is periodic
- Takens' embedding does not guarantee the point cloud will be a *nice* circle

references

[1] P. Bendich, T. Galkovskyi, and J. Harer. Improving homology estimates with random walks. *Inverse Probl.*, 2011.

[2] P. Casale, O. Pujol, and P. Radeva. Personalization and user verification in wearable systems using biometric walking patterns. *Pers. Ubi. Comp.*, 2012.

[3] R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 2006.

[4] J. Perea and J. Harer. Sliding windows and persistence: an application of topological methods to signal analysis. *Found. Comput. Math.*, 2015.

[5] F. Takens. Detecting strange attractors in turbulence. *Dyn. Sys. Turb.*, 1981.