A Fast Method of Computing Persistent Homology of Time Series Data

Shotaro Tsuji*, Kazuyuki Aihara*†

*Department of Methematical Informatics, University of Tokyo †Institute of Industrial Science, University of Tokyo

Introduction

- **Persistent homology** is applied to **reconstructed attractors** from time series data.
  - For periodicity detection and classification of time series data.
- Topological structure of orbits of ordinary differential equations reflects the periodicity or chaos.
- Computing persistent homology of thousands of points requires several days and hundreds GiB of memory.
  - We want to reduce the computational cost by fitting **cubic Bézier curves** to time series data.

Background

- Persistent homology extracts the number and the widths of holes in a shape.
- A shape is given as a simplicial complex.
- **Vietoris-Rips complex** is constructed from a point cloud.

Persistent homology

- For a point cloud X, the Vietoris-Rips complex of X is defined as

$$\mathrm{VR}_r(X) = \{|x_{i_0} x_{i_1} \cdots x_{i_q}| : q-\text{simplex} \mid \forall j,k,\ d(x_{i_j}, x_{i_k}) \le r\}.$$

- The simplices in $\mathrm{VR}_r(X)$ is sorted by the value of the filter function:

$$f(\sigma) = \begin{cases} 0, & (\dim \sigma = 0), \\ d(u,v), & (\dim \sigma = 1), \\ \max_{\tau \prec \sigma} f(\tau), & (\text{otherwise}). \end{cases}$$

The Number of Simplices in a filtration of Vietoris-Rips complex

- The threshold r is set to the infinity.
- The upper dimension of homology groups is 2, so we need simplices of dimension less than 3.
- The number of simplices is $\sum_{k=0}^{3} \binom{n}{k}$ and its order is $O(n^3)$ where n is the number of points.

Persistence Diagrams and Bottleneck Distance

- The **persistence diagram** of a persistent homology $PH_q(X)$ is defined as:

$$\mathrm{Dgm}_q(X) = \{((b(\alpha), d(\alpha)) \mid \alpha \in PH_q(X)\} \cup \{(x,x) \mid x \in \mathbb{R} \cup \{\infty\}\}.$$

- The **bottleneck distance** between diagrams X and Y is defined as:

$$W_\infty(X,Y) = \inf_{\eta: X \to Y} \sup_{\alpha \in X} \|\alpha - \eta(\alpha)\|_\infty,$$

where $\eta : X \to Y$ ranges over all bijections and the norm is defined as

$$\|\alpha\|_\infty = \|(b,d)\|_\infty = \max\{|b|, |d|\}.$$

Attractor reconstruction

- The observed time series is mapped into the **delay-coordinate space**.
- The delay-coordinate vector with dimension d and delay a is defined as

$$y(t) = (x(t), x(t-a), x(t-2a), \ldots, x(t-(d-1)a))$$

for a time series data $\{x(t)\}_{t=1}^{n}$.

Acknowledgements

This research was partially supported by AMED under Grant Number JP19dm0307009 and NEC corporation.

Proposed Method

Fit cubic Bézier curves

- The series of pairs of points and sampling time:

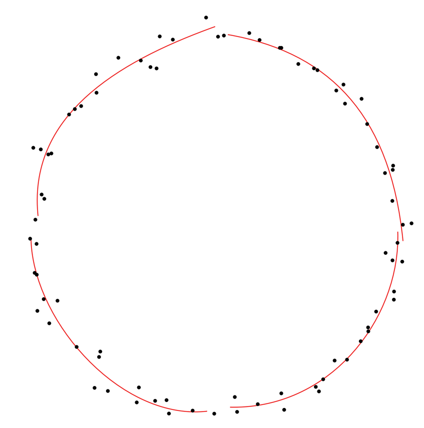
$$\{(x_i, t_i) \mid x_i \in \mathbb{R}^d\}_{i=1}^{l} \ (0 \le t_i \le 1).$$

- The input time series is grouped into several groups.
- The cubic Bézier curve is parametrized as

$$p(t) = \sum_{i=0}^{3} \binom{3}{i} (1-t)^{3-i} t^i p_i \ (0 \le t \le 1).$$

- Fitting is achieved by minimizing the squared error function:

$$L(p_0, p_1, p_2, p_3) = \sum_{i=1}^{l} \|p(t_i) - x_i\|^2.$$



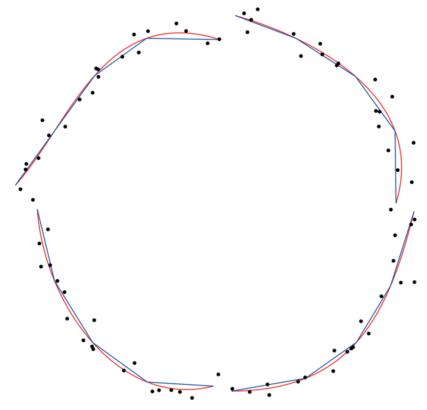
Divide into line segments

- The fitted cubic Bézier curve is divided into line segments.
- The unit interval [0, 1] is uniformly divided into r intervals:

$$[t_0, t_1], \ldots, [t_{r-1}, t_r] \text{ where } t_i = i/r \ (i = 0, 1, \ldots, r).$$

- Each interval is mapped into the cubic Bézier curve:

$$\overrightarrow{p(t_0)p(t_1)}, \overrightarrow{p(t_1)p(t_2)}, \ldots, \overrightarrow{p(t_{r-1})p(t_r)}.$$

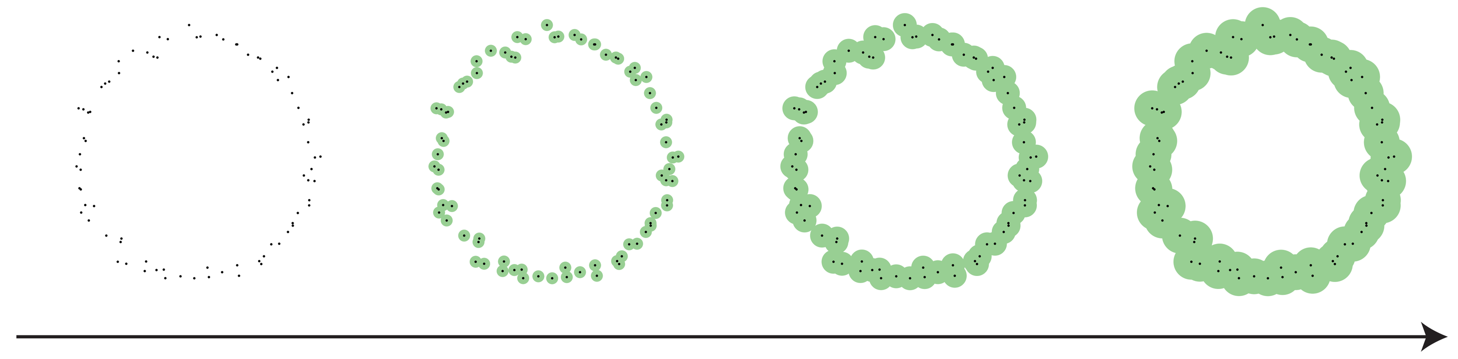


Distance between line segments

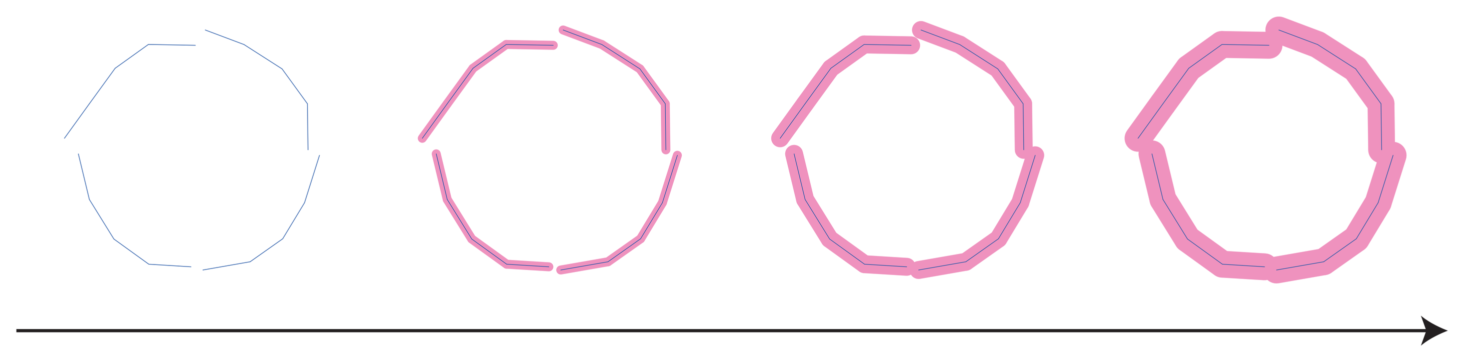
- Let $\overrightarrow{q_0 q_1}$ and $\overrightarrow{r_0 r_1}$ be line segments.
- The distance between them is defined as $d(\overrightarrow{q_0 q_1}, \overrightarrow{r_0 r_1}) = \min_{q \in \overrightarrow{q_0 q_1}, r \in \overrightarrow{r_0 r_1}} d(q, r)$.
- The distance is calculated by minimizing the function $f(s,t) = d(q(s), r(t))^2 = \|q(s) - r(t)\|^2$.

Comparison between the ordinary and proposed Method

Ordinary Method



Proposed Method



Experiments

Irrational flow on 2-torus

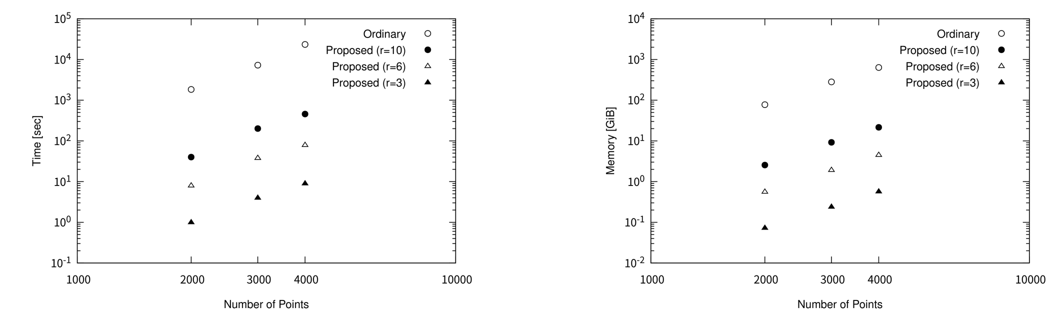
- The irrational flow is defined as $\frac{du}{dt} = \alpha \mod 1,\ \frac{dv}{dt} = \beta \mod 1$, where $(u,v) \in [0,1] \times [0,1]$.
- We set $\alpha = 1$ and $\beta = \sqrt{2}$.
- The trajectory is mapped into 3-dimensional Euclidean space with the mapping below:

$$x_1 = R\cos u + r\cos u \cos v,$$
$$x_2 = R\sin u + r\sin u \cos v,$$
$$x_3 = r\sin v,$$

where $R = 2$ and $r = 1$.

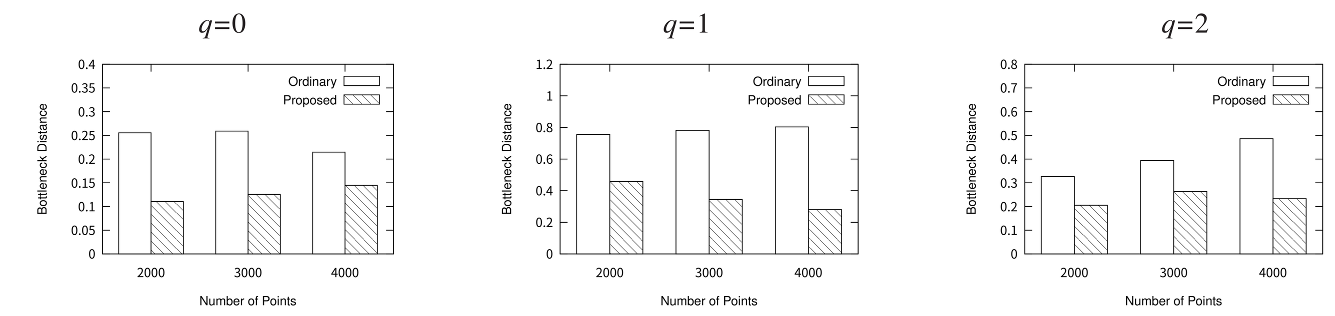
- The trajectory was developped from $t = 0$ to $t = 50\pi$.
- The number of points was n = 2000, 3000, 4000 and the parameters was l = 30 and r = 3, 6, 10.

Time performance (left) and memory performance (right)



Comparison for noised input data

- Added 10% gaussian noise to the trajectory.
- Computed the bottleneck distance between the clean input and the noised input.



Japanese vowels /a/, /e/, /i/, /o/ and /u/

- The signals were mapped into the delay-coordinate space of d = 10 and a = 10.
- 1100 steps (almost 125 ms) of each embedded signal were extracted.
- The parameter was set to l = 10 and r = 2.
- The performance of the proposed method was compared to that of the **Witness complex**.

Computational Time

Vowel	Ordinary	Proposed	Witness
/a/	290 sec	< 1 sec	< 1 sec
/e/	312 sec	< 1 sec	< 1 sec
/i/	268 sec	< 1 sec	< 1 sec
/o/	291 sec	< 1 sec	< 1 sec
/u/	277 sec	< 1 sec	< 1 sec

Computational Memory

Vowel	Ordinary	Proposed	Witness
/a/	12.5 GiB	0.11 GiB	0.14 GiB
/e/	11.5 GiB	0.09 GiB	0.09 GiB
/i/	11.5 GiB	0.09 GiB	0.10 GiB
/o/	11.5 GiB	0.09 GiB	0.09 GiB
/u/	11.5 GiB	0.09 GiB	0.09 GiB

- The distances to the ordinary Vietoris-Rips complex was compared.

Proposed Method

Vowel	0th	1st	2nd
/a/	0.064	0.644	0.131
/e/	0.029	0.032	0.047
/i/	0.026	0.027	0.014
/o/	0.021	0.068	0.011
/u/	0.016	0.032	0.020

Witness complex

Vowel	0th	1st	2nd
/a/	0.126	0.645	0.177
/e/	0.026	0.217	0.038
/i/	0.018	0.292	0.034
/o/	0.022	0.279	0.069
/u/	0.015	0.332	0.029