Summer@ICERM 2017: Topological Data Analysis
Mini-course on *Topological Time Series Analysis*

Lecturer: Jose Perea, `joperea@msu.edu`

Teaching Assistant: Christopher Tralie, `chris.tralie@gmail.com`

Introduction

Time varying observations are ubiquitous in today's data rich world; examples include real-valued time series (like sounds and temperature measurements), video data (thought of as ordered sequences of image frames) and dynamic networks (again, ordered sequences of graphs).

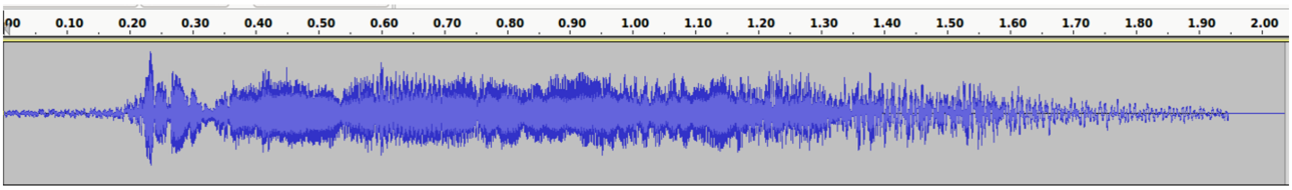


Figure 1: Sound recording of a horse whinny.

One can formalize these types of objects as functions:

Definition 1. Given a metric space $(\mathbb{M}, \mathbf{d})$, an $\mathbb{M}$-valued time series is a function

$$f : \mathbb{R} \longrightarrow \mathbb{M}$$

It follows that the sound recording above can be thought of as an $\mathbb{R}$-valued time series; if $\mathbb{M}$ is a collection $\mathbb{I}$ of images then an $\mathbb{I}$-valued time series would be a video; and if $\mathbb{M}$ is a collection $\mathbb{G}$ of graphs then a $\mathbb{G}$-valued time series is the same as a dynamic network. One of the most important tasks in time series analysis is the ability to make predictions: e.g. given the temperature now, what will be temperature in 5 minutes? One way of tackling this problem is identifying recurrent patterns in $f : \mathbb{R} \longrightarrow \mathbb{M}$. The goal of this mini-course is to show how one can use tools from algebraic topology to answer these types of questions.

Let's see a preview of how this would work. Given a positive real number τ and a positive integer d, the **sliding window embedding** of $f : \mathbb{R} \longrightarrow \mathbb{M}$ with parameters (d, τ) is the function

$$\begin{array}{rccc} SW_{d,\tau}f : & \mathbb{R} & \longrightarrow & \mathbb{M}^{d+1} \\ & t & \mapsto & \begin{bmatrix} f(t) \\ f(t+\tau) \\ f(t+2\tau) \\ \vdots \\ f(t+d\tau) \end{bmatrix} \end{array}$$

The main idea we will explore in this min-course is that if f has non-trivial recurrent structure, then the image of $SW_{d,\tau}f$ will have nontrivial topology for appropriate choices of (d, τ). Here is a warm-up exercise to convince yourself of this fact. Recall that a function $f : \mathbb{R} \longrightarrow \mathbb{M}$ is said to be periodic with period $T > 0$ if $f(t + T) = f(t)$ for all $t \in \mathbb{R}$. It turns out that periodicity of a function, which is the simplest type of recurrence, implies circularity of its sliding window embedding.

Exercise (Geometric Periodicity): For $n \in \mathbb{N}$ let $f(t) = \cos(nt)$. Show that if $d \geq 1$ and $0 < d\tau < \frac{2\pi}{n}$ then $SW_{d,\tau}f$ traces a planar ellipse in $\mathbb{R}^{d+1}$.

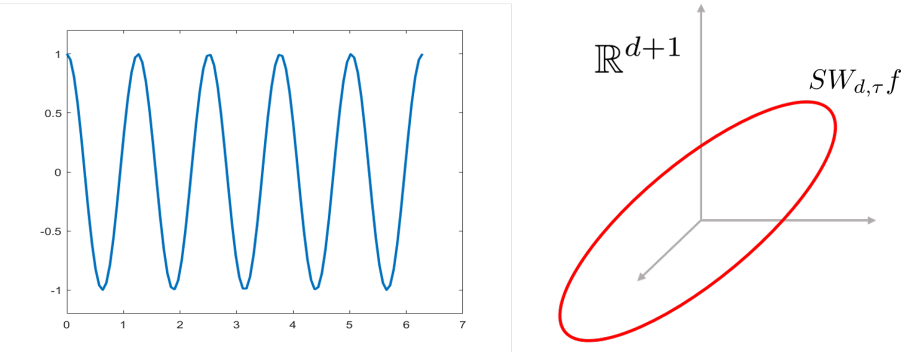
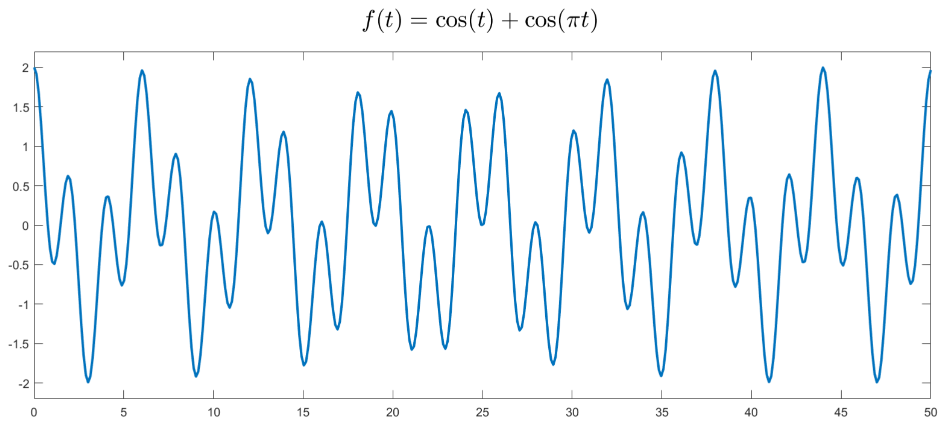


Figure 2: (Left) the function $\cos(5t)$ and (Right) a visualization of its sliding window embedding.

Here is another type of recurrent behavior: If we let $f(t) = \cos(t) + \cos(\pi t)$ then it follows that f is not periodic (**Exercise**), but its graph certainly depicts recurrent behavior.



Exercise (Non-trivial): Show that if $f(t) = \cos(t) + \cos(\pi t)$, $d \geq 4$ and $0 < d\tau < \pi$ then the set $\{SW_{d,\tau}f(k) \ : \ k \in \mathbb{Z}\}$ is dense on a 2-dimensional torus.

Theory and Applications

The main goal of the mini-course is to show how one can use computational topology, and particularly the persistent homology of sliding window embeddings, to quantify the presence of recurrent phenomena in time series data. The course will have a theoretical component, where we will further our understanding of sliding windows and persistence, as well as an applied component where we explore several questions in data science. The application areas include computational biology, music analysis and video analysis. There will be a blend of lecture style presentations, and of practical coding sections; we will provide snippets of code as well as guidance to help you make progress.

Reading Material

The mini-course will rely heavily on the material of the first week (simplicial complexes, filtrations, persistent homology and the stability theorem). Students will be expected to have a working knowledge of the persistent homology of Rips filtrations, as well as familiarity with software tools (e.g., javaPlex) to carry out explicit computations. Some useful (but not required) things to know: Every reasonable (i.e., square integrable) periodic function can be expressed as an infinite linear combination of sines and cosines (i.e. its Fourier series decomposition); Kronecker's approximation theorem, as presented here `http://www-users.math.umn.edu/~garrett/m/mfms/notes_2013-14/04_Fourier-Weyl`. The rest of the course will revolve around the references below.

References

[1] Floris Takens, *Detecting strange attractors in turbulence*, Dynamical systems and turbulence, Warwick 1980. Springer Berlin Heidelberg, 1981. 366-381.

[2] Jose Perea and John Harer. *Sliding windows and persistence: An application of topological methods to signal analysis* Foundations of Computational Mathematics 15.3 (2015): 799-838.

[3] Jose Perea et al. *Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data*, BMC Bioinformatics 16.1 (2015): 257.

[4] Jose Perea *Persistent homology of toroidal sliding window embeddings*, Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016.

[5] Christopher Tralie and Jose Perea. *(Quasi) Periodicity Quantification in Video Data, Using Topology* arXiv preprint arXiv:1704.08382 (2017).

[6] Christopher Tralie, *High-Dimensional Geometry of Sliding Window Embeddings of Periodic Videos*, LIPIcs-Leibniz International Proceedings in Informatics. Vol. 51. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.