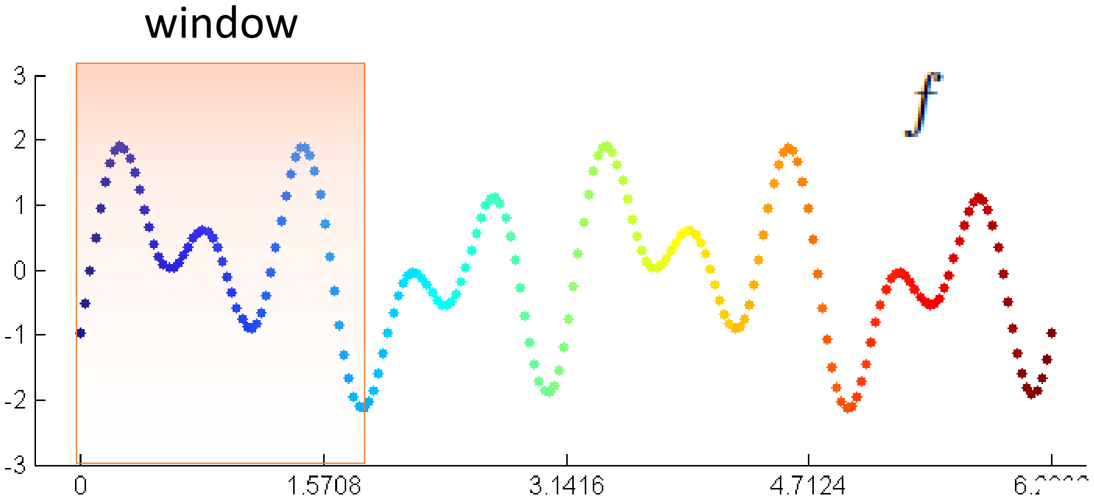
Topological Time Series Analysis

Lecture 2: Persistent Homology of Sliding Window Point Clouds

Jose Perea

Sliding window embedding

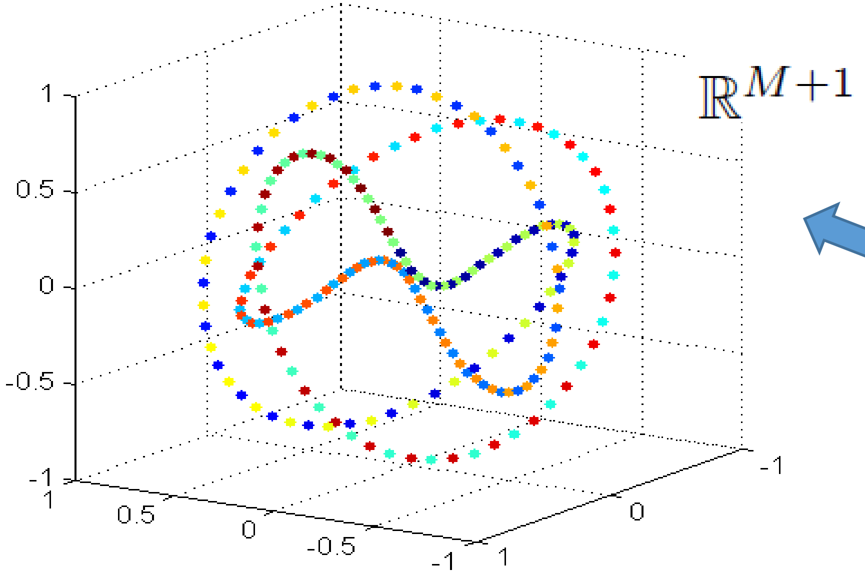


τ → Step/delay

$M\tau$ → Window size

$M+1$ → Dimension

$$SW_{M,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+M\tau) \end{bmatrix}$$

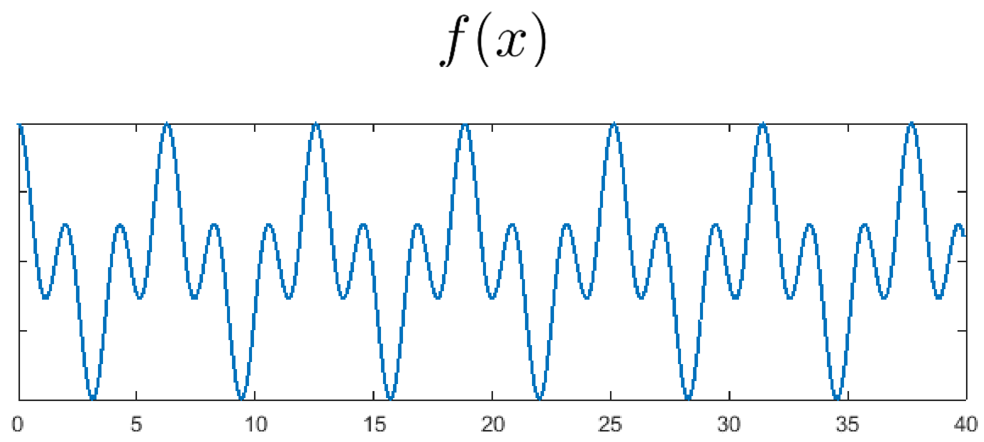


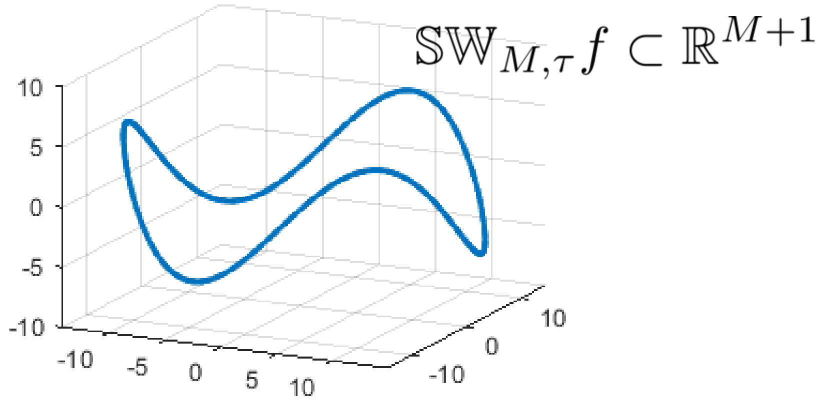
Sliding window point-cloud

$SW_{M,\tau}f(T)$

$T \subset \mathbb{R}$

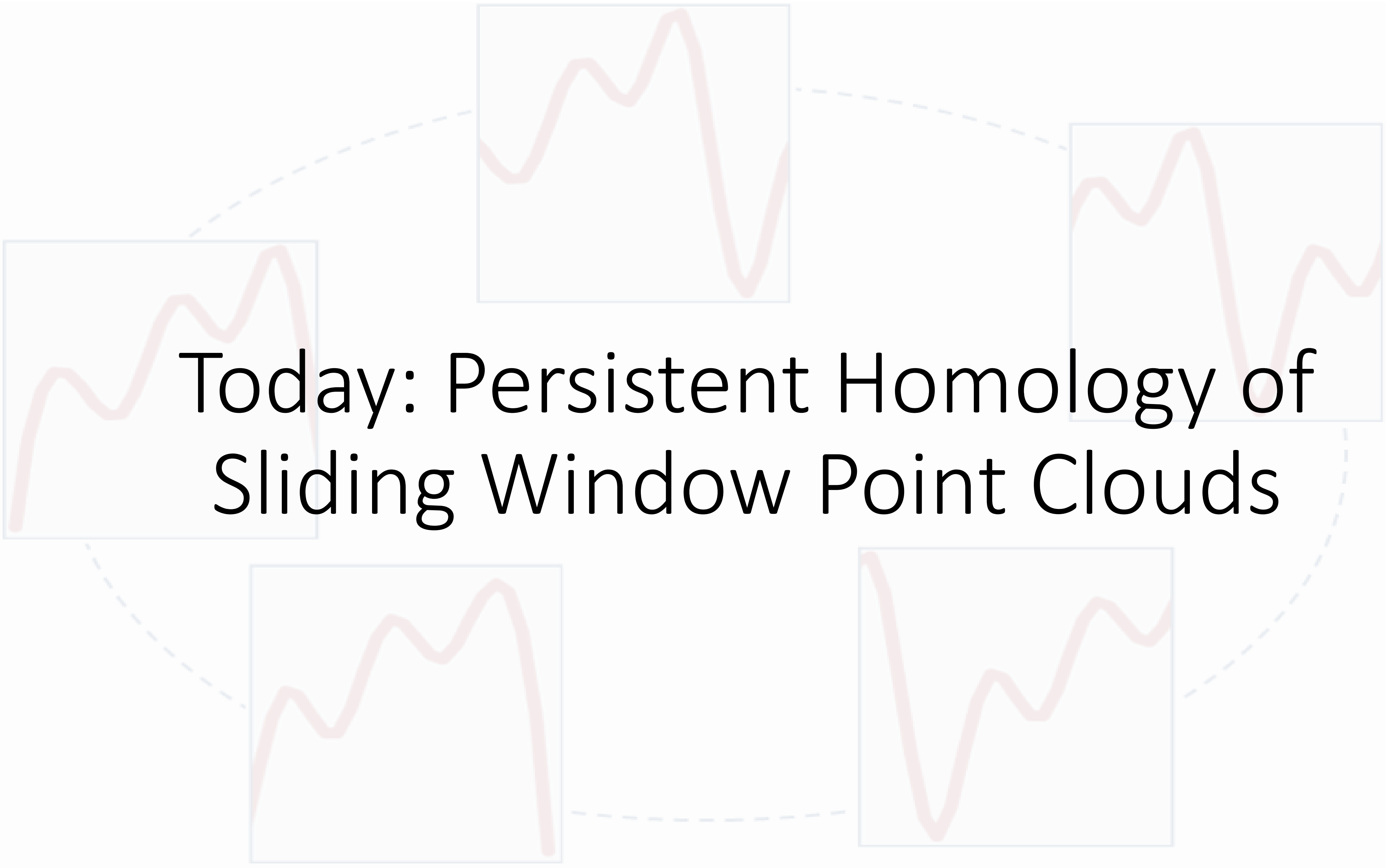
Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

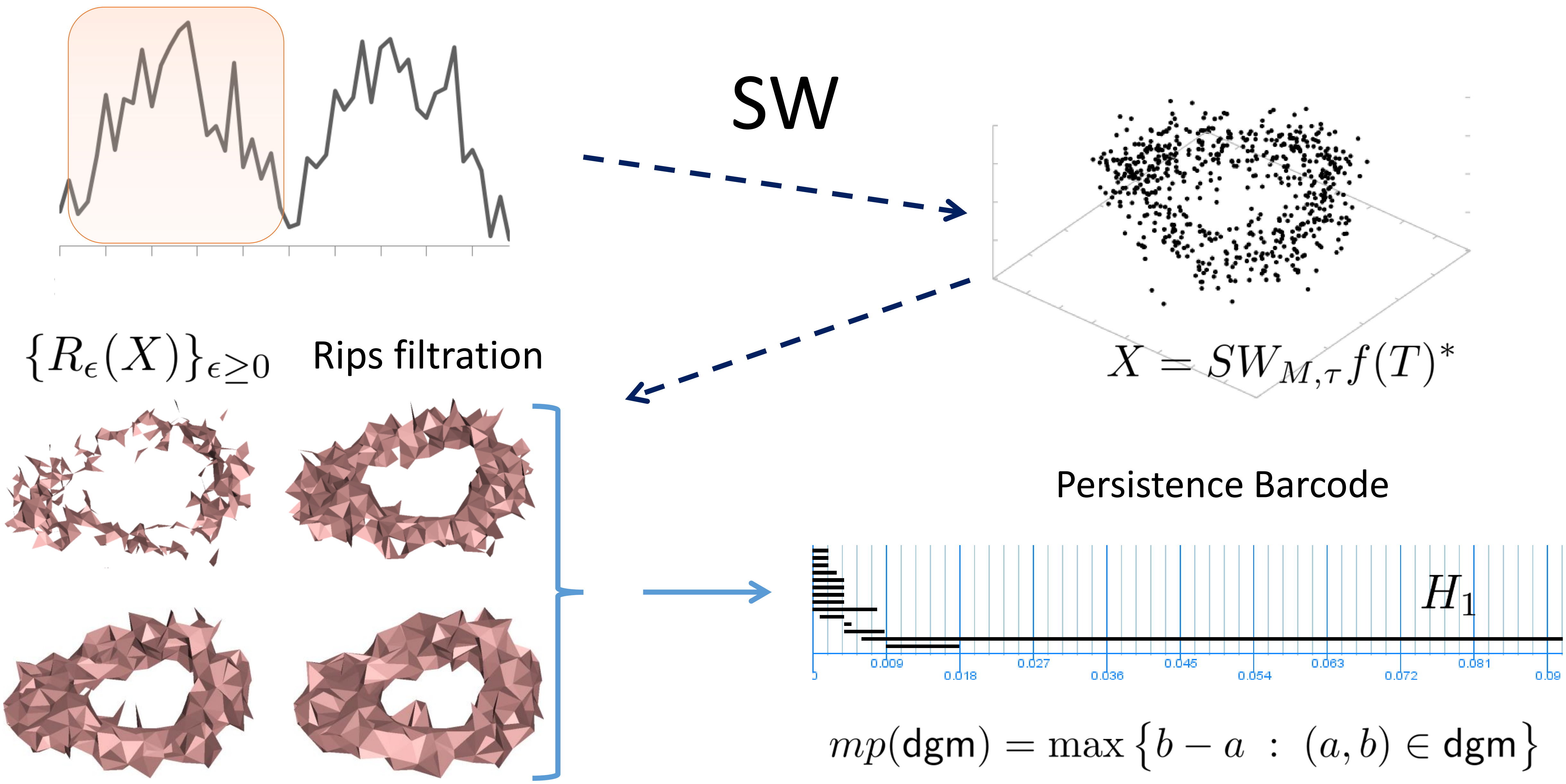




Periodicity	Circularity
Period ($f(t + 2\pi/L) = f(t)$)	Roundness (window size $M\tau = \frac{M}{M+1}\frac{2\pi}{L}$)
# of prominent harmonics (N)	Ambient Dimension ($M \geq 2N$)
# of non-commensurate frequencies	Intrinsic Dimension ($\subset S^1 \times \cdots \times S^1$)

Today: Persistent Homology of Sliding Window Point Clouds





Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

Activity 1

- Open the jupyter notebook “2-PersistentHomology”
- Is there a relation between window size and maximum persistence ?

$$M\tau$$

???

$$mp\Big(\mathsf{dgm}\big(\mathcal{R}(\mathbb{SW}_{M,\tau}f)\big)\Big) = \max\Big\{b - a \ : \ (a,b) \in \mathsf{dgm}\big(\mathcal{R}(\mathbb{SW}_{M,\tau}f)\big)\Big\}$$

Conjecture

Maximum persistence is maximized when

$$M\tau = \frac{M}{M+1}\frac{2\pi}{L}$$

Theorem (Adams, Adamaszek, 2015)

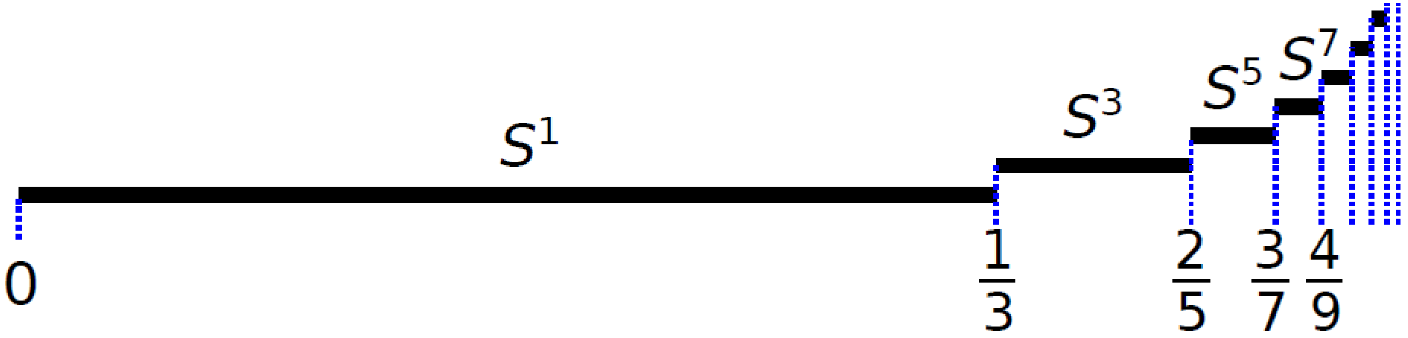
Let S^1 denote the circle of unit circumference with geodesic distance. Then

$$Rips_r(S^1) \simeq S^{2n+1}$$

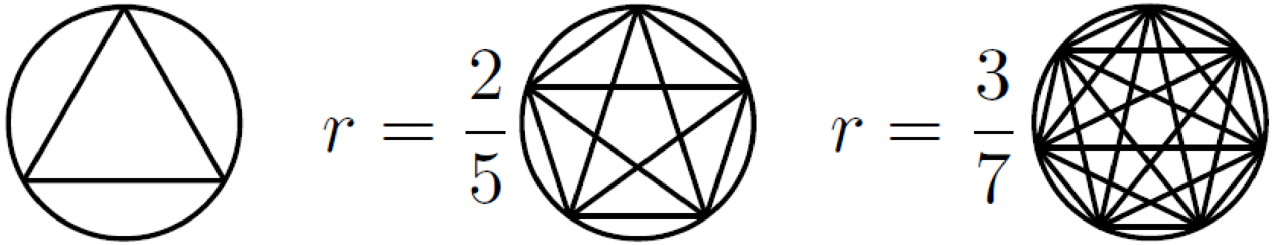
for

$$\frac{n}{2n+1} < r \leq \frac{n+1}{2n+1}$$

$$n \in \mathbb{N} \cup \{0\}$$



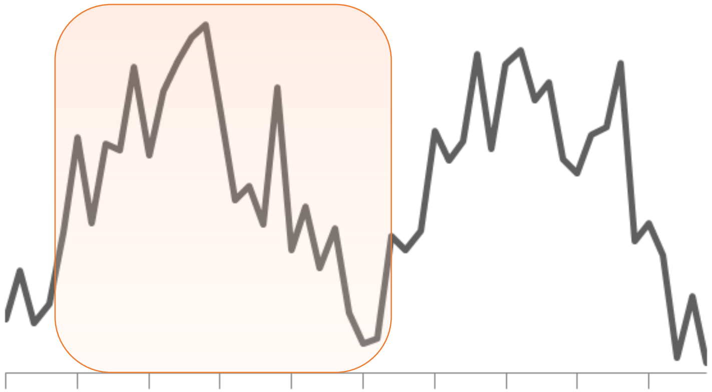
$r = \frac{1}{3}$ $r = \frac{2}{5}$ $r = \frac{3}{7}$



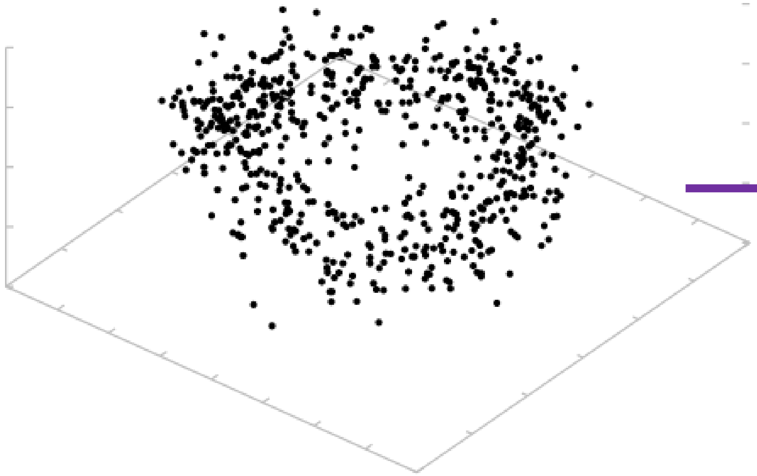
The Vietoris-Rips Complex of a Circle, H. Adams and M. Adamasezk, 2015

SW1PerS: Sliding Windows and 1-Persistence Scoring

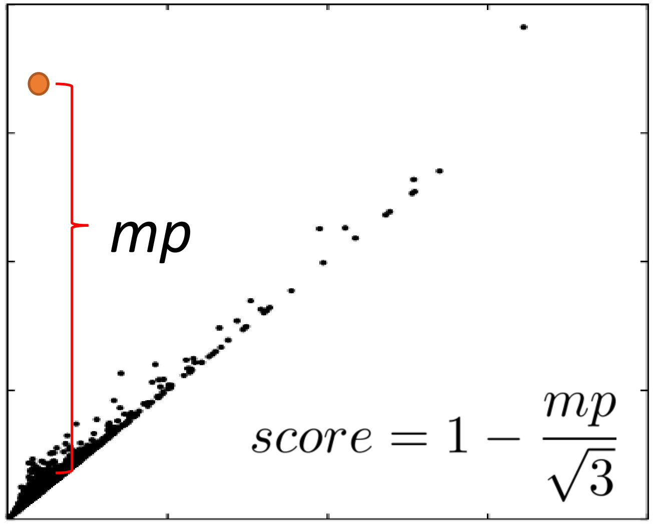
SW1PerS: Sliding Windows and 1-Persistence Scoring



$SW_{M,\tau} f(T)^*$



dgm



$N \sim$ Sample size

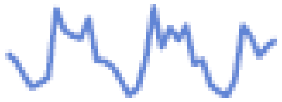
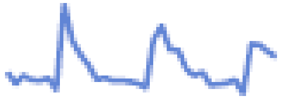
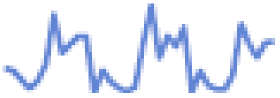
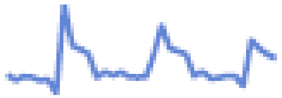
$M = 2N$

$L =$ number of periods

$$\tau = \frac{2\pi}{L(M+1)}$$

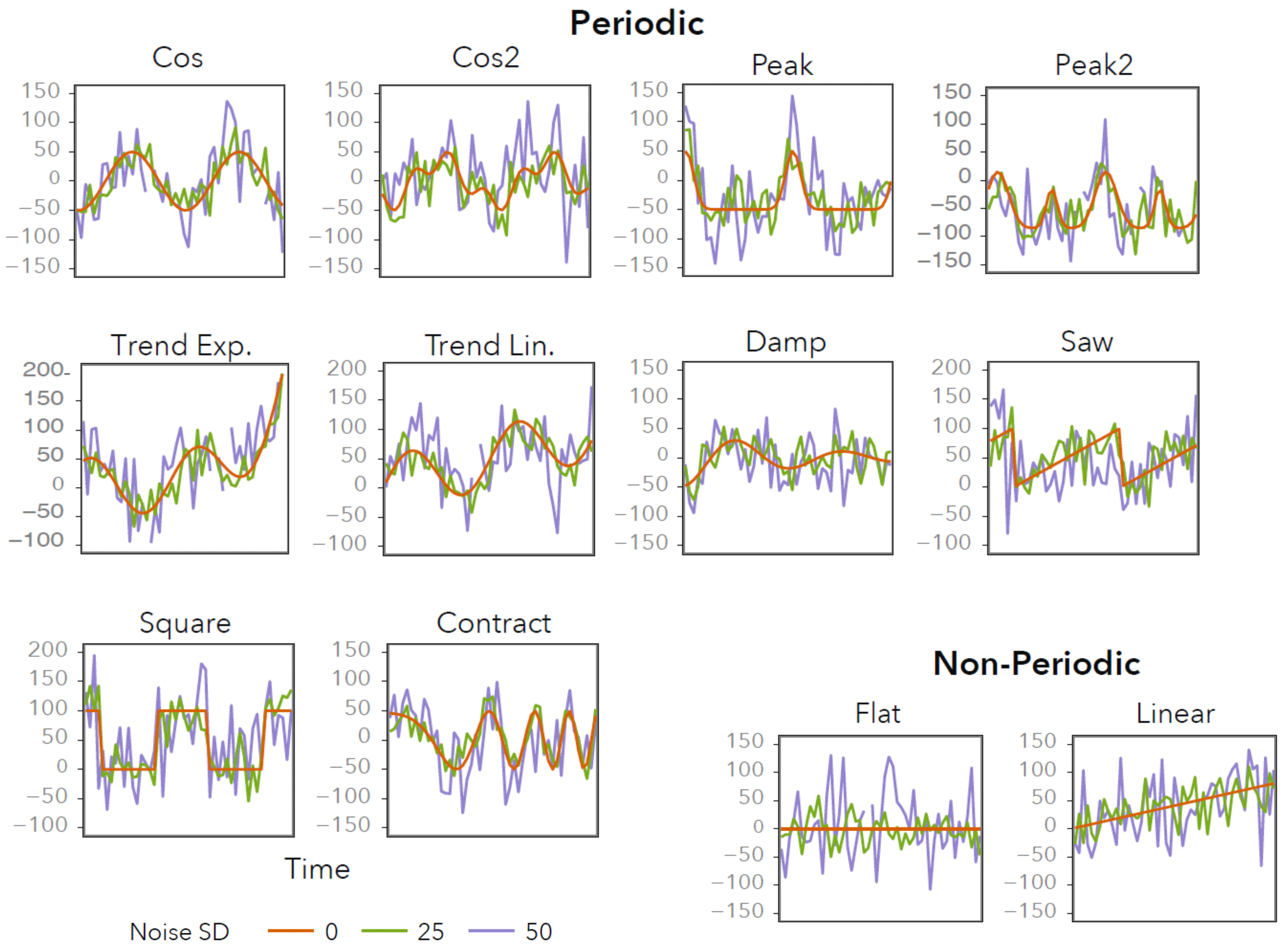
SW1PerS: Sliding Windows and 1-Persistence Scoring, J. Perea et. al., 2016

Yeast Metabolic Cycle Data

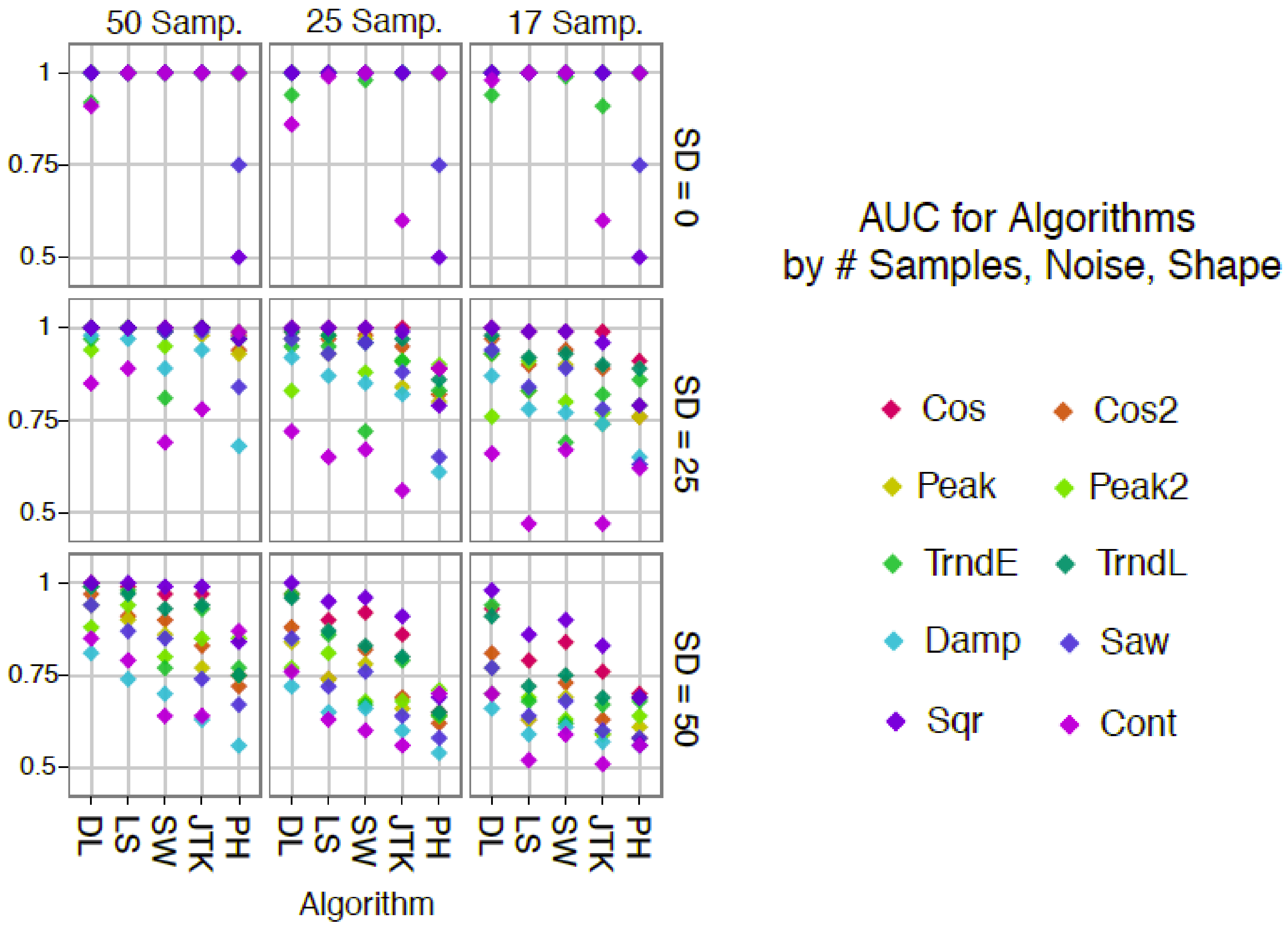
Gene	SW	DL	LS	JTK	Amp	Plot
ECM33	137	1552	1194.5	1492	35.86	
CDC9	291	1494	1993.5	2714.5	2.81	
SAM1,2	628	1133	1723	3289.5	60.82	
MSH6	715	3569	2381	3341.5	5.06	

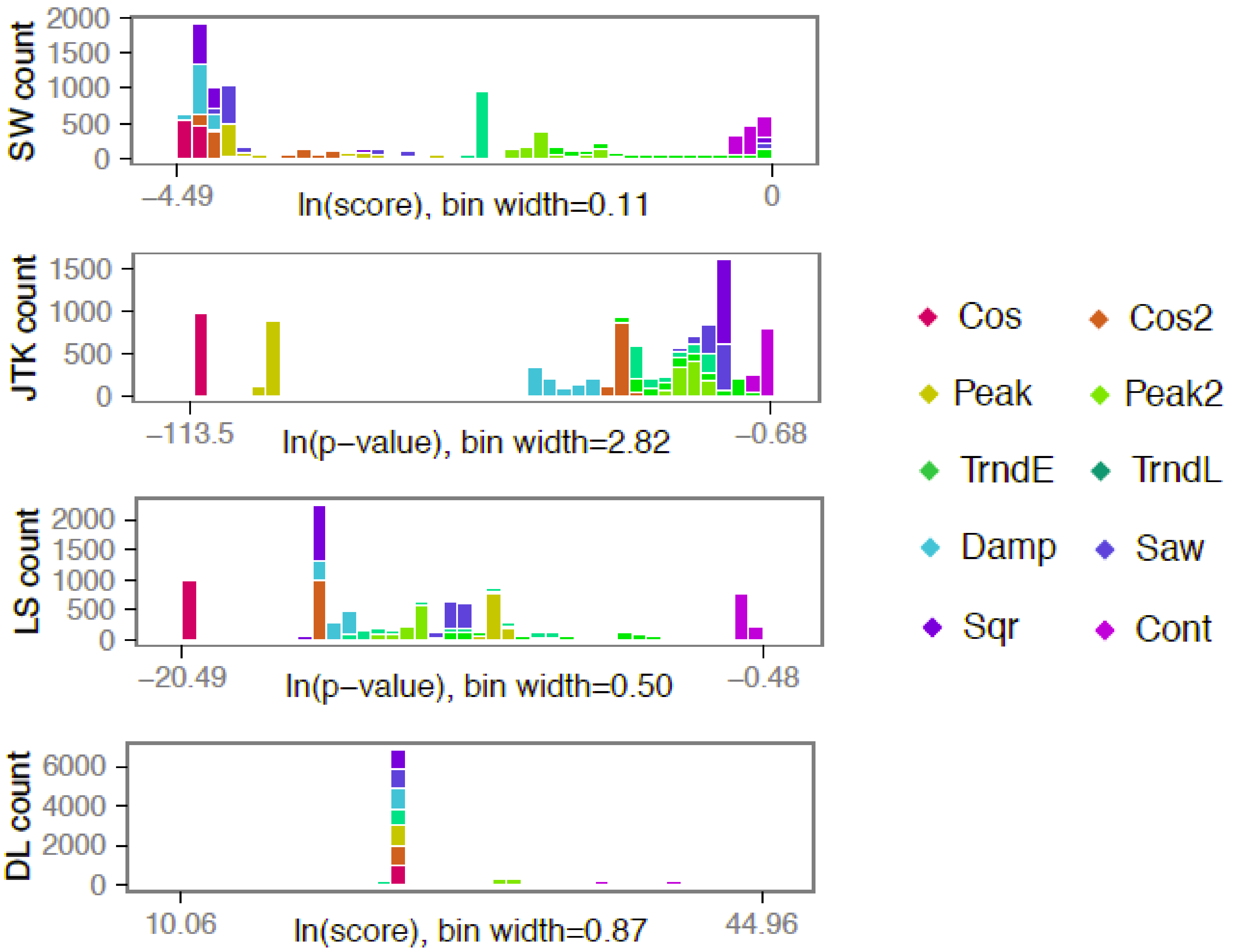
Rankings of genes in the top 10% (out of 9,330) according to SW, and not in the top 10% for any other algorithm

SW1PerS: Sliding Windows and 1-Persistence Scoring, J. Perea et. al., 2016



SW1PerS: Sliding Windows and 1-Persistence Scoring, J. Perea et. al., 2016


50 Samp.
25 Samp.
17 Samp.
1
0.75
0.5
SD = 0
SD = 25
SD = 50
DL
LS
SW
JTK
PH
Algorithm
AUC for Algorithms
by # Samples, Noise, Shape
Cos
Cos2
Peak
Peak2
TrndE
TrndL
Damp
Saw
Sqr
Cont



SW count
2000
1500
1000
500
0
−4.49
0
ln(score), bin width=0.11
JTK count
1500
1000
500
0
−113.5
−0.68
ln(p−value), bin width=2.82
LS count
2000
1500
1000
500
0
−20.49
−0.48
ln(p−value), bin width=0.50
DL count
6000
4000
2000
0
10.06
44.96
ln(score), bin width=0.87
Cos
Cos2
Peak
Peak2
TrndE
TrndL
Damp
Saw
Sqr
Cont

This page is intentionally left blank

Activity 2

- Can you differentiate sums of harmonics, from sums of non-commensurate frequencies, using persistence?

Goal:

Given $f \in L^2(\mathbb{R}/2\mathbb{Z}\,;\,\mathbb{R})$, understand

$$PH_n(\mathcal{R}(\mathbb{SW}_{M,\tau}f);\mathbb{F})$$

Persistent homology

Rips filtration

Sliding window point cloud

Field of coefficients

Strategy

- Replace $f(t)$ by its N-truncated Fourier Series

$$S_N f(t) = \sum_{n=0}^{N} a_n \cos(nt) + b_n \sin(nt)$$

- Understand the geometry of $SW_{M,\tau} S_N f(t)$

- Take the limit of the resulting 1D-diagrams as

$$N \to \infty$$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

Theorem 0 (P. and Harer)

Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, $f \in C^k(\mathbb{T},\mathbb{R})$ and let $T \subset \mathbb{T}$ be finite.

If dgm and dgm_N are the persistence diagrams of

$SW_{M,\tau} f(T)$ and $SW_{M,\tau} S_N f(T)$, respectively, then

$$d_B(\mathsf{dgm}, \mathsf{dgm}_N) \leq 2\sqrt{\frac{2}{2k-1}}\left\| f^{(k)} - S_N f^{(k)} \right\|_2 \frac{\sqrt{M+1}}{N^{k-\frac{1}{2}}}$$

Bottleneck distance

k-th derivative

L^2- metric

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

Theorem 1 (P. and Harer)

Let $f \in C^1(\mathbb{R}/2\pi\mathbb{Z}, \mathbb{R})$ be s.t. $f\left(t + \dfrac{2\pi}{L}\right) = f(t)$ for all t ($L \in \mathbb{N}$)
and so that $\|f\|_2 = 1$ and $\displaystyle\int f(t)\,dt = 0$.

1. $t \mapsto SW_{M,\tau}S_N f(t)$ is non-degenerate for $M \geq 2N$

2. $SW_{M,\tau}S_N f$ is roundest when $L(M+1)\tau = 2\pi$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

On Convergence...

Let $N \in \mathbb{N}$, $\tau_N = \dfrac{2\pi}{L(2N+1)}$, and $T \subset \mathbb{T}$

$SW_{2N,\tau_N} f(T)$ $\qquad$ $SW_{2N,\tau_N} S_N f(T)$

Pointwise mean-center
and normalize

$\mathbb{R}^{2N+1} \supset X_N$ $\qquad$ $Y_N \subset \mathbb{R}^{2N+1}$

Theorem 2 (P. and Harer).

Let $N \in \mathbb{N}$, $\tau_N = \dfrac{2\pi}{L(2N+1)}$, and $T \subset \mathbb{T}$

Then $\left\{\text{dgm}(X_N)\right\}_{N\in\mathbb{N}}$ and $\left\{\text{dgm}(Y_N)\right\}_{N\in\mathbb{N}}$

are Cauchy with respect to d_B , and

$$\lim_{N\to\infty} d_B\big(\text{dgm}(X_N), \text{dgm}(Y_N)\big) = 0$$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

Theorem 3 (P. and Harer).

Let $f \in C^1(\mathbb{R}/2\pi\mathbb{Z}, \mathbb{R})$ be s.t. $f\left(t + \frac{2\pi}{L}\right) = f(t)$ for all t ($L \in \mathbb{N}$) and so that $\|f\|_2 = 1$ and $\int f(t)\, dt = 0$.

As $M \to \infty$, with $L(M+1)\tau = 2\pi$ and $T \subset \mathbb{R}/2\pi\mathbb{Z}$ δ-dense,

then dgm $\Leftarrow SW_{\infty,0}f(T)^*$, with rational coefficients, satisfies

$$\boxed{mp(\text{dgm}) \geq 2\sqrt{3}\max_{n\in\mathbb{Z}}\left|\widehat{f}(n)\right| - 2\sqrt{2}\delta\|f'\|_2}$$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

The Field of Coefficients…

<u>Exercise</u>: Let

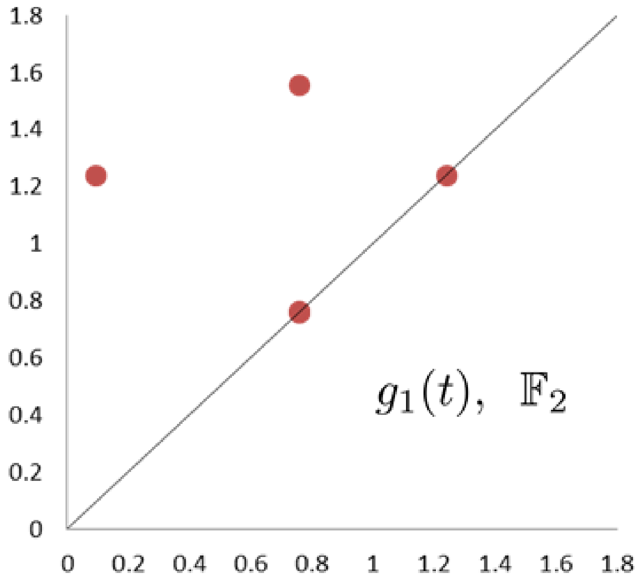
$$g_1(t) = 0.6\cos(t) + 0.8\cos(2t)$$

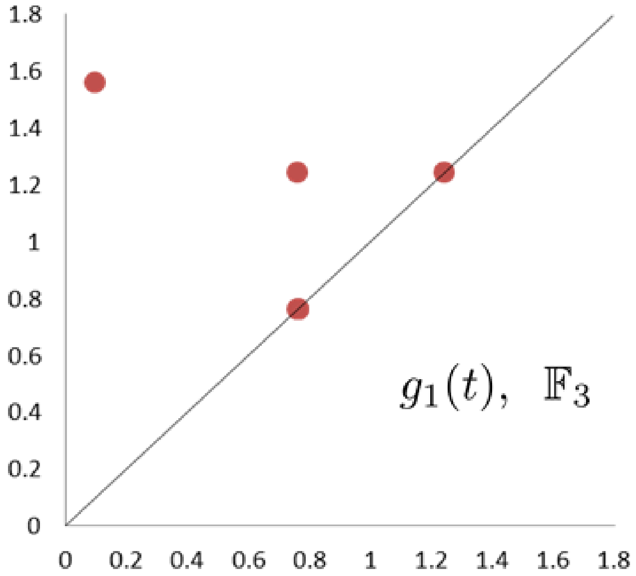
$$g_2(t) = 0.8\cos(t) + 0.6\cos(2t)$$

Compute the maximum 1D-persistence of $\mathsf{dgm}(SW_{M,\tau}g_i(T)^*)$

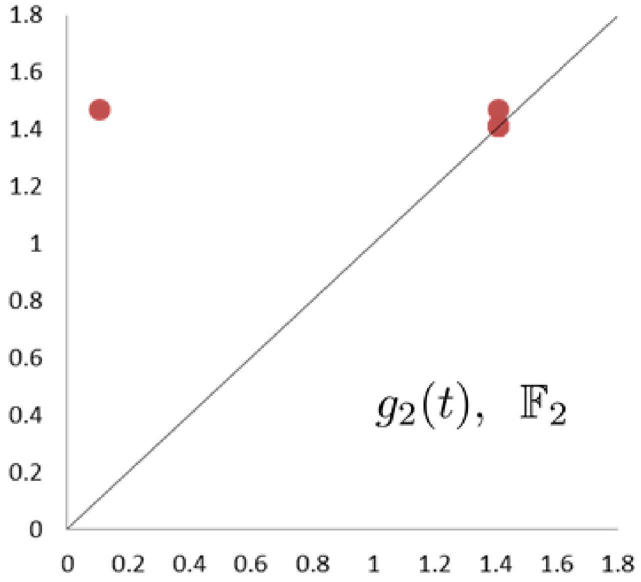
with $\mathbb{F}_2$ and $\mathbb{F}_3$ coefficients.

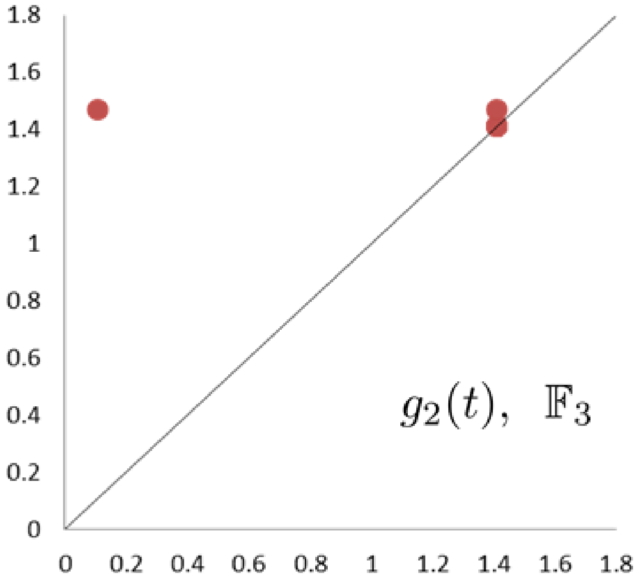
Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015





$$0.6\cos(t) + 0.8\cos(2t)$$





$$0.8\cos(t) + 0.6\cos(2t)$$

Activity 3

- Why are the persistence diagrams different?

Why does this happen…

If $g(t) = r_1 \cos(t - \alpha_1) + r_2 \cos(2t - \alpha_2)$ where

$r_1^2 + r_2^2 = 1,\ r_1 r_2 \neq 0$ and $\alpha_i \in [0, 2\pi]$, then

up to isometry

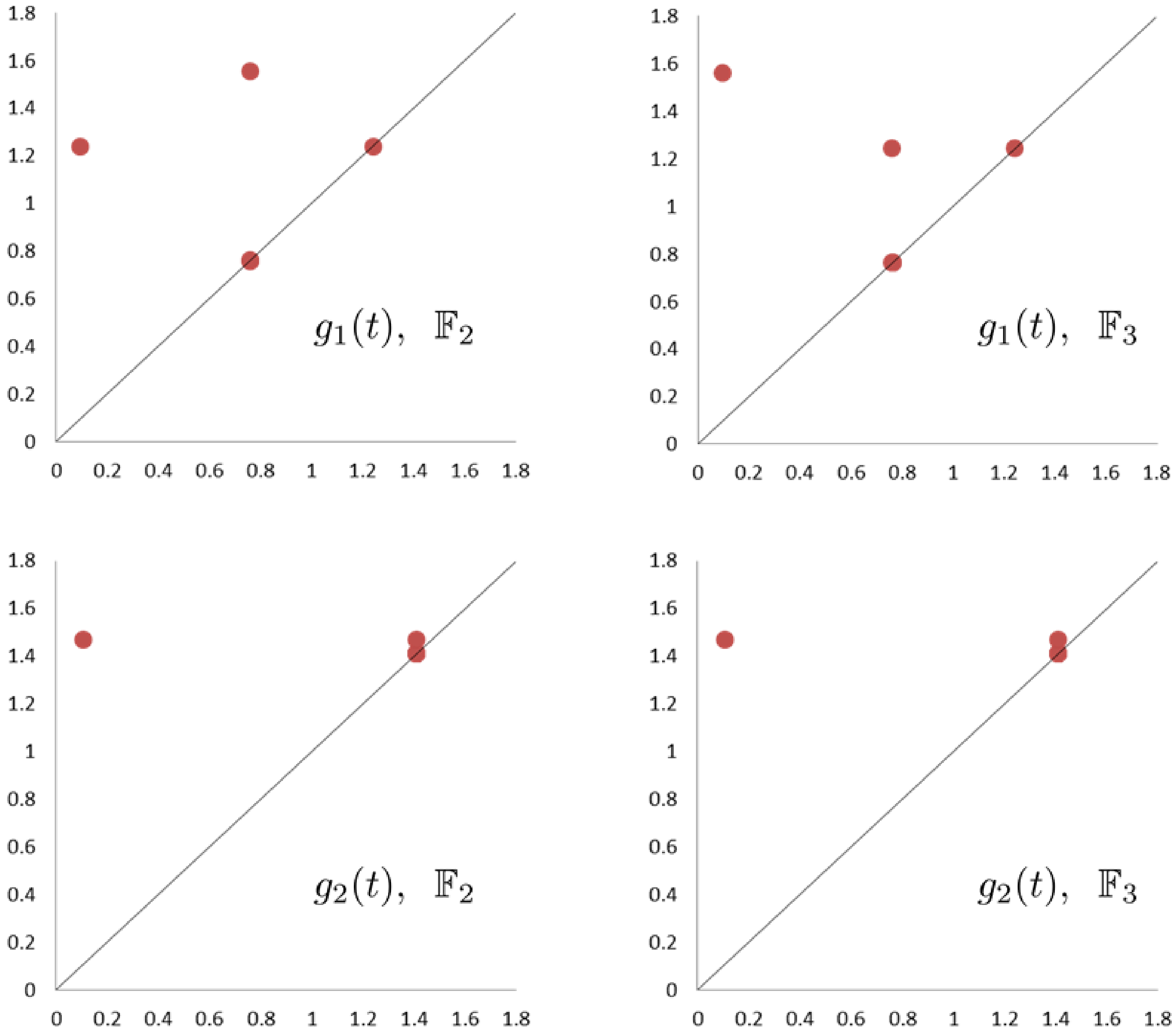
$$\boxed{SW_{M,\tau}\, g(t)^* = \left(r_1 e^{it}, r_2 e^{2it}\right)}$$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015

The bounding 2-chain for $(r_1 e^{it}, r_2 e^{2it})$

	$\mathbb{F}_2$	$\mathbb{F}_3$
$r_1 < r_2$	Mobius strip $\sim 2r_1$	Disk $\sim \sqrt{3}r_2$
$r_1 > r_2$	Disk $\sim \sqrt{3}r_1$	Disk $\sim \sqrt{3}r_1$

Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015



$$0.6\cos(t) + 0.8\cos(2t)$$

$$0.8\cos(t) + 0.6\cos(2t)$$

Thanks!

J. Perea, A. Deckard, S. Haase and J. Harer, *SW1PerS: Sliding Windows and 1-Persistence Socring; Discovering Periodicity in Time Series Data,* Preprint, 2016.

J. Perea and J. Harer, *Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis,* Foundations of Computational Mathematics, 2015.