Topological Time Series Analysis

Lecture 2: Persistent Homology of Sliding Window Point Clouds

Jose Perea
Sliding window embedding

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

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

$T \subset \mathbb{R}$

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

Period \( (f(t + 2\pi/L) = f(t)) \)

# of prominent harmonics \( (N) \)

# of non-commensurate frequencies

Circularity

Roundness \( (\text{window size } M\tau = \frac{M}{M+1} \frac{2\pi}{L}) \)

Ambient Dimension \( (M \geq 2N) \)

Intrinsic Dimension \( (\subset S^1 \times \cdots \times S^1) \)
Today: Persistent Homology of Sliding Window Point Clouds
Persistence Barcode

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

$\{R_\epsilon(X)\}_{\epsilon \geq 0}$

Rips filtration

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

Persistence Barcode

$mp(dgm) = \max \{b - a : (a, b) \in dgm\}$
Activity 1

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

\[ M_T \]

\[ m_p(\text{dgm}(\mathcal{R}(\mathcal{S}W_{M,T} f))) = \max \left\{ b - a : (a, b) \in \text{dgm}(\mathcal{R}(\mathcal{S}W_{M,T} f)) \right\} \]
Conjecture

Maximum persistence is maximized when

\[ M \tau = \frac{M}{M + 1} \cdot \frac{2\pi}{L} \]
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

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

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

---

SW1PerS: Sliding Windows and 1-Persistence Scoring
SW1PerS: Sliding Windows and 1-Persistence Scoring

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

\[ N \sim \text{Sample size} \]
\[ M = 2N \]
\[ L = \text{number of periods} \]
\[ \tau = \frac{2\pi}{L(M + 1)} \]

\[ \text{score} = 1 - \frac{mp}{\sqrt{3}} \]

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>SW</th>
<th>DL</th>
<th>LS</th>
<th>JTK</th>
<th>Amp</th>
<th>Plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECM33</td>
<td>137</td>
<td>1552</td>
<td>1194.5</td>
<td>1492</td>
<td>35.86</td>
<td></td>
</tr>
<tr>
<td>CDC9</td>
<td>291</td>
<td>1494</td>
<td>1993.5</td>
<td>2714.5</td>
<td>2.81</td>
<td></td>
</tr>
<tr>
<td>SAM1,2</td>
<td>628</td>
<td>1133</td>
<td>1723</td>
<td>3289.5</td>
<td>60.82</td>
<td></td>
</tr>
<tr>
<td>MSH6</td>
<td>715</td>
<td>3569</td>
<td>2381</td>
<td>3341.5</td>
<td>5.06</td>
<td></td>
</tr>
</tbody>
</table>

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
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

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

- 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 $\text{dgm}$ and $\text{dgm}_N$ are the persistence diagrams of $SW_{M,\tau}f(T)$ and $SW_{M,\tau}S_Nf(T)$, respectively, then

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

- 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. \( f \left( t + \frac{2\pi}{L} \right) = f(t) \) for all \( t \) (\( L \in \mathbb{N} \)) and so that \( \|f\|_2 = 1 \) and \( \int f(t) \, dt = 0. \)

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

$$SW_{2N, \tau_N} f(T) \, \xleftarrow{\text{Pointwise mean-center and normalize}}\, \xrightarrow{\text{Pointwise mean-center and normalize}} \, SW_{2N, \tau_N} S_N f(T)$$

$\mathbb{R}^{2N+1} \supset X_N \xrightarrow{\text{Pointwise mean-center and normalize}} Y_N \subset \mathbb{R}^{2N+1}$
Theorem 2 (P. and Harer).

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

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

are Cauchy with respect to $d_B$, and

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

*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 \( T \subset \mathbb{R}/2\pi\mathbb{Z} \) \( \delta \)-dense, then \( dgm \leftarrow SW_{\infty,0} f(T)^* \), with rational coefficients, satisfies

\[
mp(dgm) \geq 2\sqrt{3} \max_{n \in \mathbb{Z}} \left| \hat{f}(n) \right| - 2\sqrt{2}\delta \|f'\|_2
\]
The Field of Coefficients...

Exercise: 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 \( \text{dgm}(SW_{M, \tau} g_i(T)^*) \)

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

\[
r_1^2 + r_2^2 = 1, \quad r_1 r_2 \neq 0 \quad \text{and} \quad \alpha_i \in [0, 2\pi],
\]

then up to isometry

\[
SW_{M,\tau} g(t)^* = (r_1 e^{it}, r_2 e^{2it})
\]
The bounding 2-chain for \((r_1 e^{it}, r_2 e^{2it})\)

<table>
<thead>
<tr>
<th></th>
<th>(F_2)</th>
<th>(F_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_1 &lt; r_2)</td>
<td>Mobius strip (\sim 2r_1)</td>
<td>Disk (\sim \sqrt{3}r_2)</td>
</tr>
<tr>
<td>(r_1 &gt; r_2)</td>
<td>Disk (\sim \sqrt{3}r_1)</td>
<td>Disk (\sim \sqrt{3}r_1)</td>
</tr>
</tbody>
</table>

*Sliding Windows and Persistence: An application of topology to signal analysis, J. Perea and J. Harer, 2015*
$g_1(t), \mathbb{F}_2$

$g_1(t), \mathbb{F}_3$

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

$g_2(t), \mathbb{F}_2$

$g_2(t), \mathbb{F}_3$

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