Aplicaciones de ventanas deslizantes a periodicidad de funciones

José Miguel Quintero

November 29, 2018

Motivation

- The study of signals is very relevant in time series analysis:
  1. Early detection of diseases.
- Today: Study a method to detect periodicity and quasi-periodicity in time series signals.
- What's new? The method combines computational topology with time series analysis.

Outline

Outline

Sliding Windows

Definition

Let $f:\mathbb{R}\longrightarrow\mathbb{R}$, $M\in\mathbb{N}$ y $\tau\in\mathbb{R}$ such that $\tau>0$. Define the window of f as a point in $\mathbb{R}^{M+1}$ with base $t\in\mathbb{R}$ as

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

- The original time series $\{f(t)\}_t$ defines a point-cloud in $\mathbb{R}^{M+1}$.
- $M\tau$ is a parameter define as the window size.

Outline

Baseline example

$\cos(t)$

Let $f(t) = \cos(t)$. The the Sliding Windows of f with parameters M and τ is

$$SW_{M,\tau}f(t) = \begin{bmatrix} \cos(t) \\ \cos(t+\tau) \\ \vdots \\ cos(t+M\tau) \end{bmatrix} = \cos(t) \begin{bmatrix} 1 \\ \cos(\tau) \\ \vdots \\ \cos(M\tau) \end{bmatrix} - \sin(t) \begin{bmatrix} 0 \\ \sin(\tau) \\ \vdots \\ \sin(M\tau) \end{bmatrix}$$

Define $u = [1, \cos(\tau), \cdots, \cos(M\tau)]$ and $v = [0, \sin(\tau), \cdots, \sin(M\tau)]$. Thus, the window can be rewritten as

$$SW_{M,\tau}f(t) = \cos(t)u - \sin(t)v.$$

If u and v are linearly independent, $SW_{M,\tau}f(t)$ is a planar curve in the vector space generated by u and v.

Baseline example

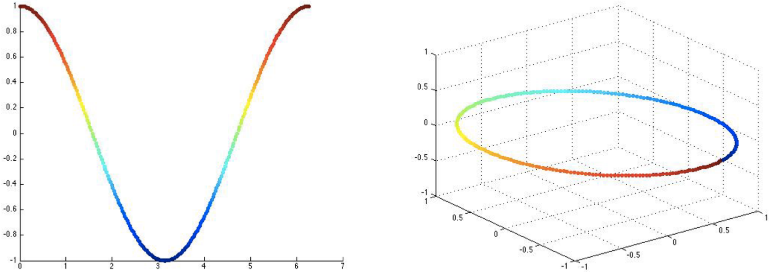
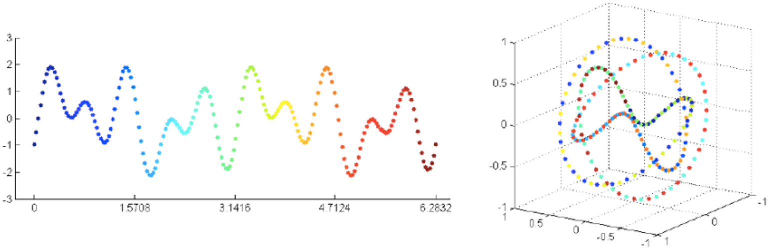


Figure: Panel to the left is plotted $f(t) = \cos(t)$ in the interval $[0, 2\pi]$. Panel to the right has $SW_{M,\tau}f(t)$ taking $M = 2$ and $\tau = \frac{2\pi}{3}$

Takeaways



- The degree to which the image of $SW_{M,\tau}f$ traces a closed curve is a measure of how periodic the function is.
- The geometry of the curve can be quite complicated to study. Thus the necessity to study the geometry of similar objects.

Outline

Simplex I

Definition

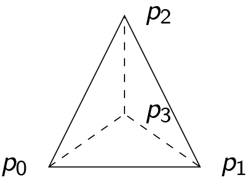
Let $P = \{p_0, p_1, \dots, p_m\}$ a set of points in $\mathbb{R}^n$. The simplex generated by P is define as

$$\Delta(P) = \left\{ \sum_{j=0}^{n} t_j p_j \middle| \; t_j \geq 0, \; \sum_{j=0}^{n} t_j = 1 \right\}.$$

Moreover, given a simplex $\Delta(P)$, for any $p_i \in P$, the opposite face of p_i is $\Delta(P \setminus \{p_i\})$.

Simplex II

Given $P = \{p_0, p_1, p_2, p_3\}$, the 3 simplex is the following



And the opposite face of p_3 is the triangle generated by $\{p_1, p_2, p_3\}$. A face is always a Simplex.

Simplicial Complex

Definition

Definition

A simplicial complex in $\mathbb{R}^n$ is an ordered pair $K = (V, \Sigma)$ where V is a set of points in $\mathbb{R}^n$ and Σ is a collection of simplex satisfying:

1. $\forall s \in \Sigma$, if $s = \Delta(P)$ then $P \subseteq V$.
2. Σ is closed under face relation.
3. If $s, b \in \Sigma$ then $s \cap b \in \Sigma$ or is empty. $s \cap b$ is define as the common face between s and b.

Moreover, the dimension of K is define as

$$\dim(K) = \sup_{s \in \Sigma} \dim(s)$$

Simplicial Complex
Example

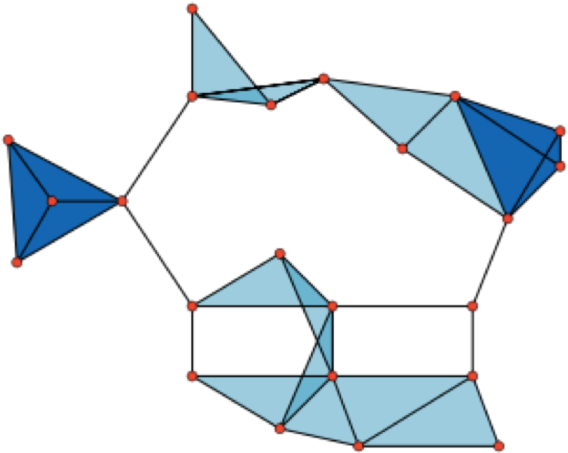


Figure: Simplicial Complex

Homology of Simplicial Complex

Definition

Let K be a simplicial complex and p a prime number. Let $\mathbb{F}_p$ be the finite field with p elements. Define $C_q(K)$ as the vector space generated by the q-dimensional simplices of K. Moreover, if $c \in C_q(K)$ then c can be expressed as

$$c = \sum_{\substack{s \in \Sigma \\ q = \dim(s)}} \gamma_s s$$

where $\gamma_s \in \mathbb{F}_p$ for all $s \in \Sigma$ such that $\dim(s) = q$.

Outline

Homology of Simplicial Complex

Definition

Let K be a simplicial complex and define $\partial_q : C_q(K) \longrightarrow C_{q-1}(K)$ as

$$\partial_q([p_0, ..., p_q]) = \sum_{i=0}^{q} (-1)^i [p_0, ..., \hat{p}_i, ..., p_q]$$

Theorem

Let K be a simplicial complex of dimension n. Then

$$0 \xrightarrow{\partial} C_n(K) \xrightarrow{\partial} C_{n-1}(K) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1(K) \xrightarrow{\partial} C_0(K) \xrightarrow{\partial} 0$$

is a chain complex.

Homology of Simplicial Complex

The past theorem allows to define the following subspaces from the vector space $C_q(K)$:

1. $Z_q(K) = \ker \partial_q$.
2. $B_q(K) = Img \partial_{q+1}$.

Definition

The q simplicial homology group of K is define as $H_q(K) = Z_q(K)/B_q(K)$ with coefficients in $\mathbb{F}_p$.

Outline

Birth and death of a homology class

Definition

Let K be a simplicial. A filtration of K is a nested sequence of subcomplex of K that starts with the empty complex and ends with K.

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K$$

- An homology class α is born at K_i if it is not in the image of the map induced by the inclusion of K_{i-1} in K_i.
- An homology class α dies entering K_j if the image of the map induced by $K_{i-1} \subset K_j$ does not contain the image of α but the image of the map induced by $K_{i-1} \subset K_{j-1}$ does.
- The persistence of α is $j - i$.

Persistence diagrams

Definition

Let K be a simplicial complex and $\{K_i\}_{i=0}^m$ be a filtration of K. Set $k \in \mathbb{N}$. The k-dimensional persistence diagram is a multiset from $\mathbb{N}^2$ where $(i, j) \in \mathrm{dgm}(k)$ if there exists an homology class associate to the $k-$homology whose birth is at K_i and dies entering K_j

- To each persistence diagram we attach the diagonal $\Delta = \{(x, x) | x \geq 0\}$ with countably multiplicity.
- The diagonal has no effect on the topological information contain the the persistence diagram since it represents classes that born and die in the same group.

Bottleneck distance

Definition

The bottleneck metric between two persistence diagrams is define as

$$d_B(\mathrm{dgm}_1(k), \mathrm{dgm}_2(k)) = \inf_{\phi} \sup_{x \in \mathrm{dgm}_1(k)} \|x - \phi(x)\|_\infty$$

where the infimum over all the bijections $\phi : \mathrm{dgm}_1(k) \longrightarrow \mathrm{dgm}_2(k)$

- ϕ always exists since the each diagram has the diagonal with countable multiplicity.

Outline

1. Introduction to Sliding Windows
   - Definition
   - Example
2. Basic Topological Concepts
   - Simplicial Complex
   - Groups of Homology
   - Persistence
   - Topological properties of the data
3. Approximating $SW_{M,\tau}f(t)$
   - Convergence of $\phi(t)$
   - Geometric Structure of the embedding
4. Empiric Exercise
   - Scoring
   - Effectiveness
5. Appendix

Rips Complex

Definition

Let S be a point cloud in $\mathbb{R}^n$ and fix $\varepsilon > 0$. Define the Rips complex of S as the simplicial comples $R(S, \varepsilon)$ with vertex set S and the simplex $\Delta(\{s_1, ..., s_k\}) \in \Sigma$ if and only if $\forall s_i, s_j \in \{s_1, ..., s_k\}$ they satisfy $\|s_i - s_j\| < \varepsilon$.

- Since $R(S, \varepsilon_1) \subseteq R(S, \varepsilon_2)$ whenever $\varepsilon_1 \leq \varepsilon_2$ then a filtration of the full complex is induced through the Rips complex.
- For small distances the simplicial complex derived is completely disconnected where each point is an individual simplex.

Rips Complexes

Example

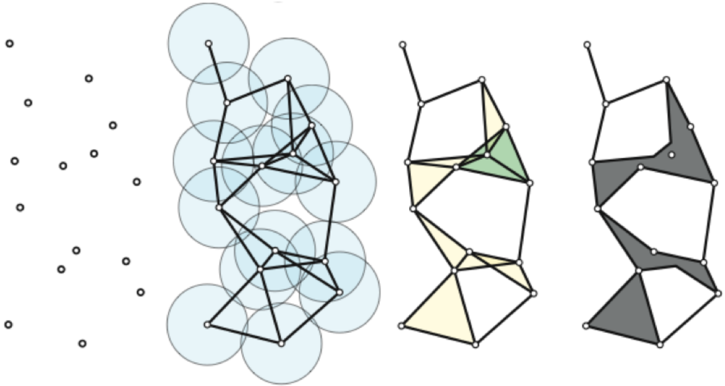


Figure: Construction of the Rips Complex for a fix ϵ

Rips Complexes

Example

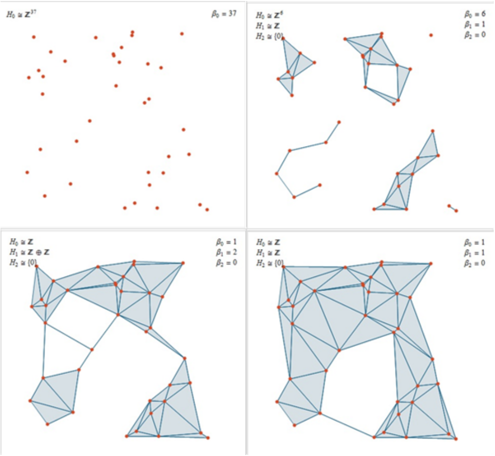


Figure: Filtration derived from Rips Complex

Stability

Proposition

Let X and Y be two point clouds. The persistent diagrams associated to each point cloud is stable. This means

$$d_b(\mathrm{dgm}(X), \mathrm{dgm}(Y)) \leq 2d_H(X, Y)$$

where d_H denotes the Hausdorff distance define as:

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$$

Computational Cost

- Studying the geometry of objects generated by $SW_{M,\tau}f$ can be difficult and computational expensive.
- It is easier to study trigonometric polynomials and thus the Fourier approximation of f.
- In this section we show that $SW_{M,\tau}$ behaves well under Fourier approximation.
- Moreover, this approximations work in the context of persistent diagrams

Fourier Approximation

The Fourier expansion of the Sliding Window of f

$$SW_{M,\tau}f(t) = \sum_{n=0}^{N} \cos(nt)(a_n u_n + b_n v_n) + \sin(nt)(b_n u_n - a_n v_n) + SW_{M,\tau}R_N f(t)$$

where $u_n = SW_{M,\tau}\cos(nt)\big|_{t=0}$, $v_n = SW_{M,\tau}\sin(nt)\big|_{t=0}$, a_n and b_n are the coefficients of the Fourier expansion. To ease the notation let

$$\phi(t) = \sum_{n=0}^{N} \cos(nt)(a_n u_n + b_n v_n) + \sin(nt)(b_n u_n - a_n v_n)$$

Outline

$SW_{M,\tau}$ as a linear operator

Theorem

$\forall M \in \mathbb{N}$ and $\tau > 0$, $SW_{M,\tau} : C(\mathbb{T}, \mathbb{R}) \longrightarrow C(\mathbb{T}, \mathbb{R}^{M+1})$ is a linear bounded operator and its bound is $\|SW_{M,\tau}\| \leq \sqrt{M+1}$.

Proof

Uniform Convergence

Theorem

Fix $k \in \mathbb{N}$. For $f \in C^k(\mathbb{T}, \mathbb{R})$ then for each $t \in \mathbb{T}$ the following inequality holds:

$$\| SW_{M,\tau} f(t) - \phi_\tau(t) \|_{\mathbb{R}^{M+1}} \leq \frac{\sqrt{2(M+1)}}{N^{k-1/2}\sqrt{2k-1}} \left\| R_N f^{(k)} \right\|_2$$

Max. Persistence

Definition

Let $(x, y) \in \text{dmg}$, and define $\text{pers}(x, y) = y - x$. The max. persistence of the diagram is

$$mp(\text{dmg}) = \max_{x \in \text{dmg}} \text{pers}(x, y)$$

Approximation Theorem

Theorem

Let $T \subseteq \mathbb{T}$, $f \in C^k(\mathbb{T},\mathbb{R})$, $X = SW_{M,\tau}f(T)$ *y* $Y = \phi(T)$. *Then*

1. $d_H(X,Y) \leq \dfrac{\sqrt{2(M+1)}}{N^{k-1/2}\sqrt{2k-1}} \left\| R_N f^{(k)} \right\|_2$
2. $\left| mp(\mathrm{dgm}(X)) - mp(\mathrm{dgm}(Y)) \right| \leq 2d_B\left(\mathrm{dgm}(X), \mathrm{dgm}(Y)\right)$
3. $d_B\left(\mathrm{dgm}(X), \mathrm{dgm}(Y)\right) \leq \dfrac{2\sqrt{2(M+1)}}{N^{k-1/2}\sqrt{2k-1}} \left\| R_N f^{(k)} \right\|_2$

Outline

Construction of the Window

- So far we have studied the topological properties of the window and the Fourier approximation.
- How to choose M and τ?
- Trade off for M:
  1. $M+1$ is the detail of the function. Higher values of M yields better results.
  2. Computational expensive.
- Trigonometric polynomials: no information is lost whenever the dimension of the embedding is greater than twice the maximum frequency.
  - $S_N f$ can be recovered from $SW_{M,\tau}f$ if $u_0, u_1, v_1, \ldots, u_N, v_N$ are linearly independent.

No loss of Information

Proposition

Fix $M\tau < 2\pi$. Then $u_o, u_1, v_1, ..., u_N, v_N$ are linearly independent $M \geq 2N$.

Assumption: From now on, given $n \in \mathbb{N}$, set $M = 2N$ and choose $\tau > 0$ so that $M\tau < 2\pi$.
Fundamental relation between window size, $1D$ persistence, and underlying frequency: the maximum persistence of the sliding-window point cloud from $S_N f$ is largest when the window size $M\tau$ is proportional to the underlying frequency.

Fourier Coefficients

Proposition

Foreach $n \geq 1$, $\langle u_n, v_n \rangle = \|u_n\|^2 - \|v_n\|^2 = 0$ *if and only if*

$$n(M+1)\tau \equiv 0 \pmod{\pi}.$$

The last proposition implies also that $a_n u_n + b_n v_n$ is orthogonal to $b_n u_n - a_n v_n$ for any $a_n, b_n \in \mathbb{R}$

Proposition

Let f *be an* L*-periodic function and assume that* $L(M+1)\tau = 2\pi$. *Then the set of vector*

$$\{u_n, v_n | 0 \leq n \leq N, \; n \equiv 0 \pmod{L}\}$$

is orthogonal and has norm $\|u_n\| = \|v_n\| = \sqrt{\frac{M+1}{2}}$ *para* $n \equiv_L 0$.

Centering Theorem

Theorem

Let $C : \mathbb{R}^{M+1} \longrightarrow \mathbb{R}^{M+1}$ be the centering function. If f an L-periodic function, $L(M+1)\tau = 2\pi$ then:

1. $\phi_\tau = \hat{f}(0)\mathbf{1} + C(\phi_\tau(t))$
2. *There is an orthogonal set of vectors* $\left\{\tilde{x}_n, \tilde{y}_n \in \mathbb{R}^{M+1} \;\middle|\; 1 \leq n \leq N \;\; n \equiv 0(mod\ L)\right\}$ *such that*

$$\varphi_\tau(t) = \frac{C(\phi_\tau(t))}{\|C(\phi_\tau(t))\|} = \sum_{\substack{n=1 \\ n\equiv 0\ (modL)}}^{N} \tilde{r}_n(\cos(nt)\tilde{x}_n + \sin(nt)\tilde{y}_n)$$

where $\tilde{r}_n = \dfrac{2\left|\hat{f}(n)\right|}{\sqrt{\|S_N f\|_2^2 - \hat{f}(0)^2}}$.

Outline

Lower Bound

Theorem

Let $f \in C^1(\mathbb{T})$ be a L-periodic function satisfying $\hat{f}(0 = 0)$ and $\|f\|_2 = 1$. Let $T \subseteq \mathbb{T}$ be a finite set such that $d_H(T, \mathbb{T}) < \delta$ with

$$0 < \delta < \frac{\sqrt{3}}{\sqrt{2}\|f'\|_2} \max_{n \in \mathbb{N}} \left|\hat{f}(n)\right|$$

Then if H_1 is a $\mathbb{Q}$ vectorial space, the 1D persistence diagram $\mathrm{dgm}_\infty(f, T, w)$ *satisfies*

$$\frac{1}{2} mp\left(\mathrm{dgm}_\infty(f, T, w)\right) \geq \sqrt{3} \max_{n \in \mathbb{N}} \left|\hat{f}(n)\right| - \sqrt{2}\delta\|f'\|_2$$

and thus

$$mp\left(\mathrm{dgm}_\infty(f, w)\right) \geq 2\sqrt{3} \max_{n \in \mathbb{N}} \left|\hat{f}(n)\right|$$

Scoring

1. Let $S = [s_1, s_2, ..., s_J]$ be a sampled signal.
2. Estimate f_S using a cubic spline interpolation.
3. Construct X_S the point cloud from the center and normalized sliding window.
4. Define its periodicity score as

$$\text{Score}(S) = \frac{mp(\text{dmg}(X_S))}{\sqrt{3}}$$

Outline

Time Series

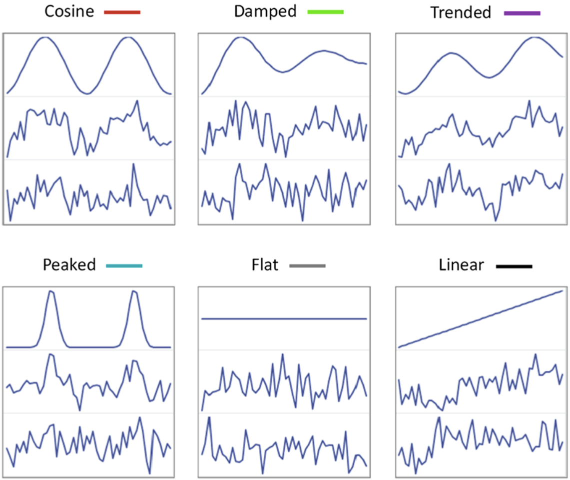
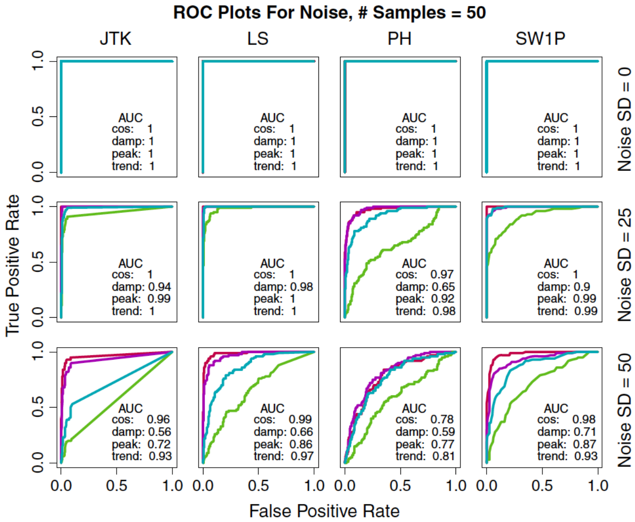


Figure: Profiles with three Gaussian noise levels.

ROC Curves


ROC Plots For Noise, # Samples = 50
JTK
LS
PH
SW1P
Noise SD = 0
Noise SD = 25
Noise SD = 50
True Positive Rate
False Positive Rate
AUC cos: 1 damp: 1 peak: 1 trend: 1
AUC cos: 1 damp: 1 peak: 1 trend: 1
AUC cos: 1 damp: 1 peak: 1 trend: 1
AUC cos: 1 damp: 1 peak: 1 trend: 1
AUC cos: 1 damp: 0.94 peak: 0.99 trend: 1
AUC cos: 1 damp: 0.98 peak: 1 trend: 1
AUC cos: 0.97 damp: 0.65 peak: 0.92 trend: 0.98
AUC cos: 1 damp: 0.9 peak: 0.99 trend: 0.99
AUC cos: 0.96 damp: 0.56 peak: 0.72 trend: 0.93
AUC cos: 0.99 damp: 0.66 peak: 0.86 trend: 0.97
AUC cos: 0.78 damp: 0.59 peak: 0.77 trend: 0.81
AUC cos: 0.98 damp: 0.71 peak: 0.87 trend: 0.93
0.0
0.5
1.0

Thank You

Proof
Bounded Linear Operator

Linearity is trivial. It is left to show that it is bounded. Let $f \in L^2(\mathbb{T})$ and consider the norm of $\mathbb{R}^{M+1}$ for $t \in \mathbb{T}$ fix. Then:

$$\begin{aligned}
\|SW_{M,\tau}f(t)\|^2_{\mathbb{R}^{M+1}} &= \sum_{n=0}^{M} |f(t+n\tau)|^2 \\
&\leq \sum_{n=0}^{M} \sup_{t\in\mathbb{T}} |f(t)|^2 \\
&= \sum_{n=0}^{M} \|f\|^2 \\
&= (M+1)\|f\|^2
\end{aligned}$$

Return