Topological summaries of periodic-like functions

Wojciech Reise (wojciech.reise@inria.fr)

ongoing work with F. Chazal, B. Michel.

Séminaire de Mathématiques appliquées
Feb 7, 2023, Nantes

TopAI chair
ANR-19-CHIA-0001

Magnetic signal for vehicle navigation

Advanced navigation systems estimate the position of a vehicle by aggregating estimates from different sensors:

- GPS,
- inertial sensors (accelerometer, gyrometer).

Adding position or movement information based on measurements from other, independent sensors can lead to an improvement in the resulting estimation.

Magnetic signal for vehicle navigation

Advanced navigation systems estimate the position of a vehicle by aggregating estimates from different sensors:

- GPS,
- inertial sensors (accelerometer, gyrometer).

Adding position or movement information based on measurements from other, independent sensors can lead to an improvement in the resulting estimation.

The magnetic field measured inside a moving car is the Earths' field perturbed by quantities

+ related to the movement of the car
  - the heading of the vehicle,
  - the rotations of the wheels,
  - the revolutions of the engine,

- those independent thereof
  - passing vehicles,
  - high–voltage installations,
  - infrastructure.

Model of the magnetic field

Let $S : [0, T] \to \mathbb{R}^3$ be the magnetic field measured by a sensor inside a vehicle. Assuming that the Earths' magnetic field is constant (locally),

$$S(t) = \psi_{\theta(t)}(\gamma(t)) \; + \; W(t), \tag{1}$$

where

1. ψ_θ is the (periodic) perturbation induced by the position of the wheels γ,
2. θ is the heading,
3. W represents noise (sensor noise, passing vehicle or electric infrastructure).

[1] Thomas Bonis et al. (May 2022). "Topological Phase Estimation Method for Reparameterized Periodic Functions". In: DOI: 10.48550/arXiv.2205.14390.

Model of the magnetic field

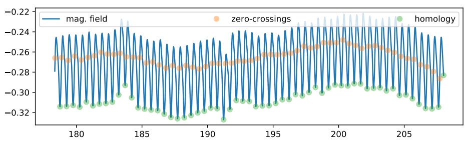
Let $S : [0, T] \to \mathbb{R}^3$ be the magnetic field measured by a sensor inside a vehicle. Assuming that the Earths' magnetic field is constant (locally),

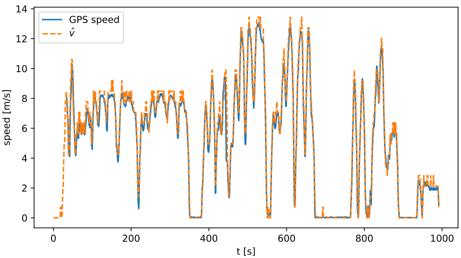
$$S(t) = \psi_{\theta(t)}(\gamma(t)) \; + \; W(t), \tag{1}$$

where

1. ψ_θ is the (periodic) perturbation induced by the position of the wheels γ,
2. θ is the heading,
3. W represents noise (sensor noise, passing vehicle or electric infrastructure).

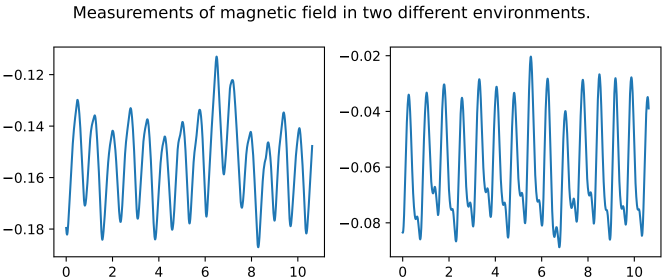
We can estimate γ by studying the periodic structure of S. In[1], we developed a method to count the number of oscillations.

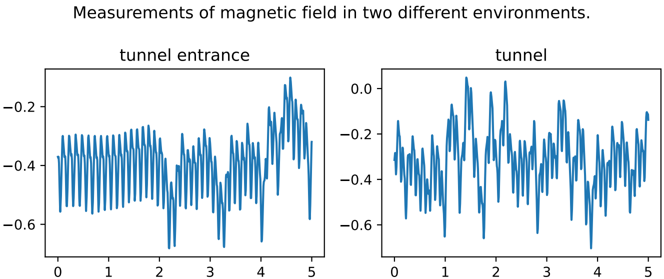




[1] Thomas Bonis et al. (May 2022). "Topological Phase Estimation Method for Reparameterized Periodic Functions". In: DOI: 10.48550/arXiv.2205.14390.

The periodic function depends on the environment





Topological signatures

Properties of the limit representation

Invariance to reparametrisation

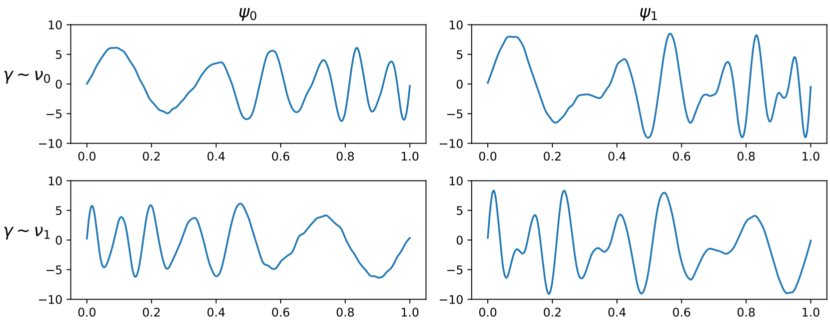
Estimation

Numerical illustration

Mathematical problem statement

Consider $S = \psi \circ \gamma + W$, where

- $\psi : \mathbb{R} \to \mathbb{R}$ is 1–periodic,
- $\gamma : [0, T] \to [0, R]$ an increasing bijection (random), $\gamma \sim \nu$,
- $W : [0, T] \to \mathbb{R}$ is a continuous stochastic process, $W \sim \mu$.



Aim

Construct a signature of ψ from S.
Test for $\psi_1 = \psi_2$, based on observations S_1, S_2, where S_k is as above, with $W_k \sim \mu$ and $\gamma_k \sim \nu_k$.

Prior work and context

The problem of comparing (populations of) curves, up to reparametrisation and constructing their representations is tackled shape analysis through methods of two types:

1. find reparametrisations, which align curves and then do standard statistics (mostly, calculate means)[2],
2. Frechet mean for a specific metric, Square Root Velocity (SRV)[3].

The models present limitations

1. Both methods are relevant when the signal has the same length, for example: growth curves, migration of birds.
2. The phase variations are "small".

In addition, the object of interest is a curve representative of the population of curves.

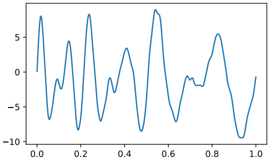
Our idea is based on topological summaries

1. Statistics on prominent local extrema - no need to know how many cycles we observe.
2. Generic asymptotic results for independent and dependent data.
3. In contrast to standard methods on time series, it is invariant to reparametrisation.

[2] J. S. Marron et al. (Nov. 2015). "Functional Data Analysis of Amplitude and Phase Variation". In: *Statistical Science* 30.4, pp. 468–484. ISSN: 0883-4237. DOI: 10.1214/15-STS524. arXiv: 1512.03216.

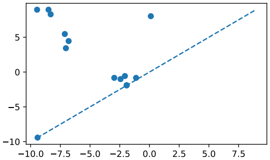
[3] A Srivastava et al. (July 2011). "Shape Analysis of Elastic Curves in Euclidean Spaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.7, pp. 1415–1428. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.184.

Topological signatures



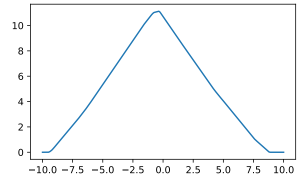
S
Observed signal

$\longrightarrow$



$D(S)$
Persistence diagram

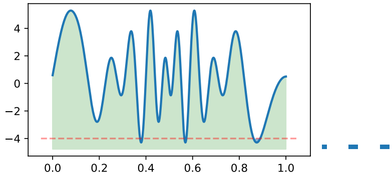
$\longrightarrow$



$\bar{\rho}(D(S))$
Functional representation

Persistence diagram of sublevel sets

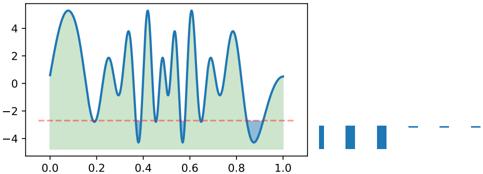
The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].



[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Persistence diagram of sublevel sets

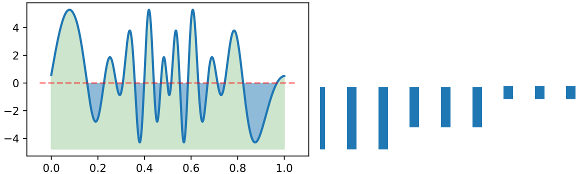
The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].



[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Persistence diagram of sublevel sets

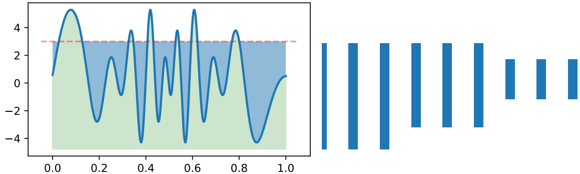
The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].



[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Persistence diagram of sublevel sets

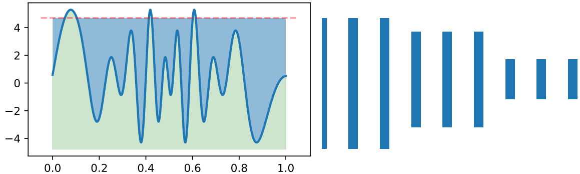
The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].



[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Persistence diagram of sublevel sets

The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].



[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Persistence diagram of sublevel sets

The persistence diagram $D(S)$ of sub level–sets of a continuous function $S : [0, T] \to \mathbb{R}$ is a point measure defined on the upper half–plane Δ_+ above $\Delta := \{x = y\}$[4].

The persistence diagram captures the height and order of local extrema.

[4] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Properties of the persistence diagram

Proposition (Invariance to reparametrisation)

For any two increasing bijections $\gamma_1, \gamma_2 : [0, T] \to [0, R]$,

$$D(\psi \circ \gamma_1) = D(\psi \circ \gamma_2). \tag{2}$$

In addition, there exists $c \in [0, 1]$, *such that for any* $N \in \mathbb{N}$,

$$D(\psi|_{[c,c+N]}) = ND(\psi|_{[c,c+1]}). \tag{3}$$

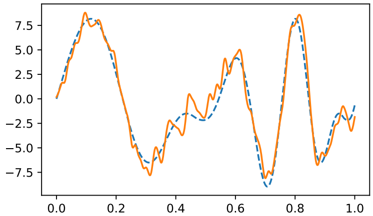
$\Longrightarrow$ the order and height of extrema characterize the persistence diagram

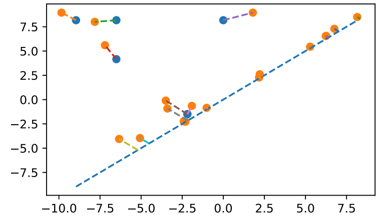
Properties of the persistence diagram: stability

Proposition (Stability Theorem[5])

For any $W : [0, T] \to \mathbb{R}$,

$$d_B(D(\psi \circ \gamma + W), D(\psi \circ \gamma)) \leq \|W\|_\infty . \tag{4}$$





The persistence of a point (b, d) is $w(b, d) = d - b$. The greater its persistence, the more prominent the extrema that generated it.

[5] Frédéric Chazal et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Normalized functional representations

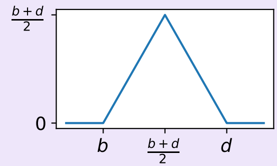
The space of multisets lacks mathematical structure[6]. It is common to use vectorisations[7], especially in a statistical learning context.

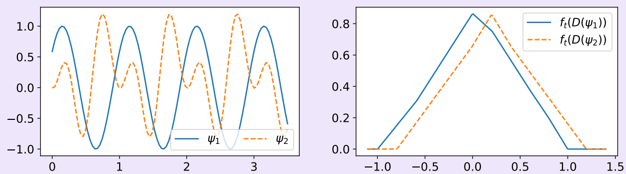
Persistence Silhouette[8]

The persistence silhouette of D is

$$\bar{\rho}_{\cdot}(D) : t \mapsto \frac{\sum_{x \in D} w_{\epsilon}(x)^p \Lambda_x(t)}{\sum_{x \in D} w_{\epsilon}(x)^p}, \tag{5}$$

where $\Lambda_{(b,d)}(t) = \max(0, \min(t-b, d-t))$ and $w_{\epsilon}(b,d) = \max(d-b-\epsilon, 0)$.





[6]Henry Adams and Michael Moy (May 2021). "Topology Applied to Machine Learning: From Global to Local". In: *Frontiers in Artificial Intelligence* 4, p. 668302. ISSN: 2624-8212. DOI: 10.3389/frai.2021.668302.

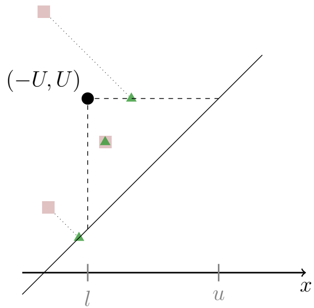
[7]Mathieu Carrière et al. (Mar. 2020). "PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures". In: *arXiv:1904.09378 [cs, math, stat]*. arXiv: 1904.09378 [cs, math, stat]; Henry Adams et al. (Jan. 2017). "Persistence Images: A Stable Vector Representation of Persistent Homology". en. In: *The Journal of Machine Learning Research* 18.1, pp. 218–252; Peter Bubenik (Jan. 2015). "Statistical Topological Data Analysis using Persistence Landscapes". en. In: *Journal of Machine Learning Research* 6, pp. 77–102.

[8]Frédéric Chazal et al. (2014). "Stochastic Convergence of Persistence Landscapes and Silhouettes". en. In: *Annual Symposium on Computational Geometry - SOCG'14*. Kyoto, Japan, pp. 474–483. (Visited on 03/05/2021).

Projections of diagrams

For $U > 0$, let $\pi_U : \Delta_{\geq 0} \to \Delta_{\geq 0}$ be the operator which projects points above the diagonal, onto the upper half-square with corner at $(-U, U)$

$$\begin{array}{rccl} \pi_U : & \Delta_{\geq 0} & \to & \Delta_{\geq 0} \\ & (x, y) & \mapsto & (x, y) + (1, -1)\min(\max(y - U, -U - x, 0), \frac{y-x}{2}). \end{array} \tag{6}$$



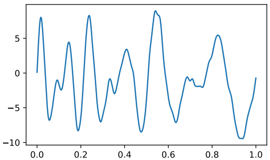
Silhouettes of projected diagrams

Fix $U > 0$ and set $\bar{\rho}_t^U := \bar{\rho}_t \circ \pi_U$. Then, $\mathcal{F} = (\bar{\rho}_t^U)_{t \in [-U,U]}$

1. has bounded support,
2. is Lipschitz with respect to t (uniformly in D)

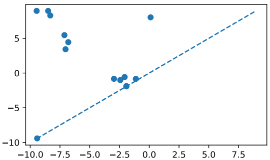
$$\forall D,\ |\bar{\rho}_{t_1}^U(D) - \bar{\rho}_{t_2}^U(D)| \leq |t_1 - t_2|,\ \forall t_1, t_2. \tag{7}$$

Topological signature



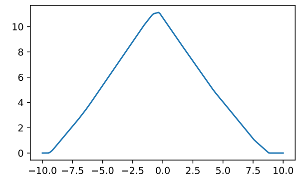
S
Observed signal

$\longrightarrow$



$D(S)$
Persistence diagram

$\longrightarrow$



$\bar{\rho}(D(S))$
Functional representation

Finally, we define

$$F(S) : t \mapsto \mathbb{E}[\bar{\rho}_t(\pi_u(D(S)))]. \tag{8}$$

Limit in the number of observed periods

Theorem

There exists $c \in [0, 1]$ such that

$$\left\| \bar{\rho}_t(D(\psi|_{[0,R]})) - \bar{\rho}_t(D(\psi|_{[c,c+1]})) \right\|_\infty \xrightarrow[R\to\infty]{} 0. \tag{9}$$

Limit in the number of observed periods

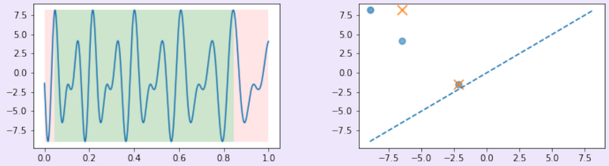
Theorem

There exists $c \in [0,1]$ such that

$$\left\| \bar{\rho}_t(D(\psi|_{[0,R]})) - \bar{\rho}_t(D(\psi|_{[c,c+1]})) \right\|_\infty \xrightarrow[R\to\infty]{} 0. \quad (9)$$

Idea behind the proof

When $W = 0$, the $D(\psi|_{[0,R]}) = \lfloor R-2 \rfloor D^* + D'$.



In addition, $\mathrm{pers}^p_{p,\epsilon}(D) = \lfloor R-2 \rfloor \mathrm{pers}^p_{p,\epsilon}(D^*) + \mathrm{pers}^p_{p,\epsilon}(D')$, so

$$\bar{\rho}(D(\psi|_{[0,R]})) = \bar{\rho}(D^*) + O\left(\frac{1}{\lfloor R-2 \rfloor}\right).$$

Since we normalize, by the total number of oscillations we observe an increasing number of periods, the signature converges.

Limit in the number of observed periods

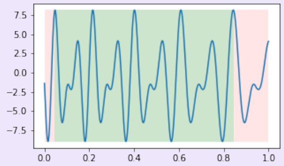
Theorem

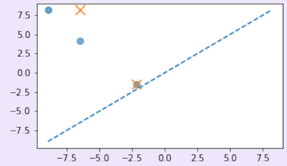
There exists $c \in [0,1]$ such that

$$\left\| \bar{\rho}_t(D(\psi|_{[0,R]})) - \bar{\rho}_t(D(\psi|_{[c,c+1]})) \right\|_\infty \xrightarrow[R\to\infty]{} 0. \tag{9}$$

Idea behind the proof

When $W = 0$, the $D(\psi|_{[0,R]}) = \lfloor R-2 \rfloor D^* + D'$.





In addition, $\mathrm{pers}^p_{p,\epsilon}(D) = \lfloor R-2 \rfloor \mathrm{pers}^p_{p,\epsilon}(D^*) + \mathrm{pers}^p_{p,\epsilon}(D')$, so

$$\bar{\rho}(D(\psi|_{[0,R]})) = \bar{\rho}(D^*) + O\left(\frac{1}{\lfloor R-2 \rfloor}\right).$$

Since we normalize, by the total number of oscillations we observe an increasing number of periods, the signature converges.

Justifies the terminology *the signature of* ψ.

Discriminating periodic functions

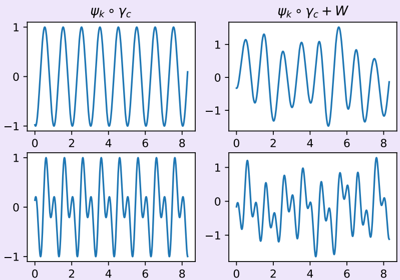
Linear transformations

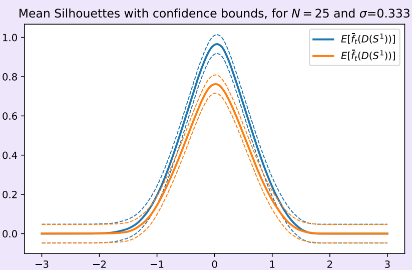
Let $\psi_2 = a\psi_1 + \theta$, for some $a > 0$ and $\theta \in \mathbb{R}$. Then,

$$\bar{\rho}_t(D(\psi_2)) = a\bar{\rho}_{(t-\theta)/a}(D(\psi_1)). \tag{10}$$

Small bumps

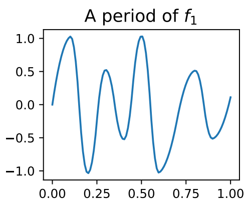
Suppose that $T = 8.5$ and that $\gamma(t) = t + c$, for $c \sim \mathcal{U}([0,1])$. $S = \psi \circ \gamma + W$, where W has covariance $(s,t) \mapsto \sigma^2 \exp(-\frac{(s-t)^2}{2\ell^2})$, with $\ell = 0.3$ and $\sigma = 1/3$.

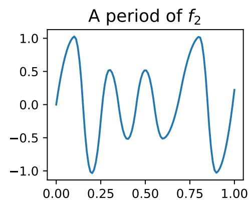




Confusing certain functions

Certain different periodic functions have the same persistence diagram and so the same functional.

D (

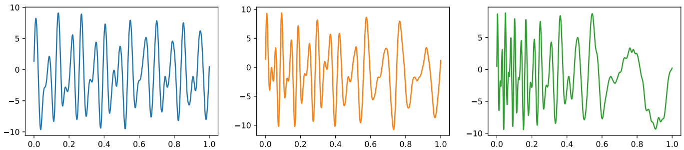
) $= D$ (

)

In particular, f_1 looks like periods of a function, and it has the same diagram as f_2.

Invariance to γ

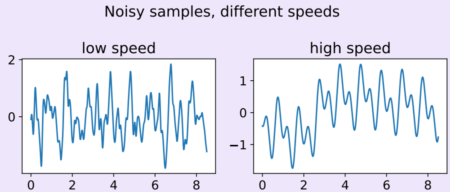
The parametrisation $\gamma : [0, T] \to [0, R]$ represents a trajectory, from $x_0 = 0$ to $x_1 = R$. We would like $\bar{\rho}$ to be the same, regardless of the trajectory from x_0 to x_1.

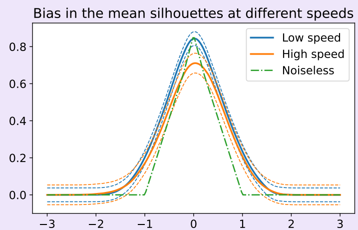


It is the case when $W = 0$, by Proposition 1.

Bias in the noisy case

In the extreme case of high noise, different time-scales lead to a bias





Stability with respect to the distribution of γ

Noise

We assume that W is

1. has Hölder–continuous paths,
2. uniformly bounded by $(\max \psi - \min \psi - \epsilon)/2$.

Reparametrisations

Consider $\gamma_1 \sim \nu_1$ and $\gamma_2 \sim \nu_2$. We assume

1. same, **fixed** endpoints: $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(T) = \gamma_2(T)$.
2. that there exists $v_{\min} > 0$, such that $v_{\min}|t - s| \leq |\gamma(t) - \gamma(s)|$, for $\gamma = \gamma_1, \gamma_2$.

Consider $\gamma_1 \sim \nu_1$ and $\gamma_2 \sim \nu_k$.

Theorem

Under the assumptions above, the signature is Hölder–continuous with respect to μ

$$\|F(\psi \circ \gamma_1 + W) - F(\psi \circ \gamma_2 + W)\| \leq \frac{L_k \left(1 + 4pUC_{\psi,\mu}\right) C_W}{v_{\min}} W_{1,\|\cdot\|_\infty}(\nu_1, \nu_2)^\alpha. \quad (11)$$

Stability with respect to the distribution of γ

Noise

We assume that W is

1. has Hölder–continuous paths,
2. uniformly bounded by $(\max \psi - \min \psi - \epsilon)/2$.

Reparametrisations

Consider $\gamma_1 \sim \nu_1$ and $\gamma_2 \sim \nu_2$. We assume

1. same, **fixed** endpoints: $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(T) = \gamma_2(T)$.
2. that there exists $v_{\min} > 0$, such that $v_{\min}|t-s| \leq |\gamma(t) - \gamma(s)|$, for $\gamma = \gamma_1, \gamma_2$.

Consider $\gamma_1 \sim \nu_1$ and $\gamma_2 \sim \nu_k$.

Theorem

Under the assumptions above, the signature is Hölder–continuous with respect to μ

$$\|F(\psi \circ \gamma_1 + W) - F(\psi \circ \gamma_2 + W)\| \leq \frac{L_k\left(1 + 4pUC_{\psi,\mu}\right) C_W}{v_{\min}} W_{1,\|\cdot\|_\infty}(\nu_1, \nu_2)^\alpha. \tag{11}$$

$C_{\psi,\mu} = \max_k \frac{pers_{p-1,\epsilon}^{p-1}(\psi\circ\gamma+W_k)}{pers_{p,\epsilon}^{p}(\psi\circ\gamma+W_k)}$, $C_W = \mathbb{E}[Poly(\Lambda_W)]$, *with* $Poly(0) = 0$.

Estimation of the signature from time series

In practice, we observe a time series $(S_n)_{n=1}^{N}$,

$$S_n = \psi(\gamma(t_n)) + W(t_n). \tag{12}$$

Can we estimate a signature of the process from these observations?

Let X be a window of length M from S. We estimate $F(X)$ using the empirical mean $\hat{F}$, calculated on blocks $X_1, \dots X_{N-M+1}$ from S

$$X_n = (S_n, \dots S_{n+M-1}).$$

1. $F(X)$ and $F(S)$ are similar, if $W = 0$, by Theorem 1.
2. Is $\hat{F}$ a good approximation of $F(X)$?

Estimation of the signature from time series

In practice, we observe a time series $(S_n)_{n=1}^N$,

$$S_n = \psi(\gamma(t_n)) + W(t_n). \tag{12}$$

Can we estimate a signature of the process from these observations?

Let X be a window of length M from S. We estimate $F(X)$ using the empirical mean $\hat{F}$, calculated on blocks $X_1, \dots X_{N-M+1}$ from S

$$X_n = (S_n, \dots S_{n+M-1}).$$

1. $F(X)$ and $F(S)$ are similar, if $W = 0$, by Theorem 1.
2. Is $\hat{F}$ a good approximation of $F(X)$?

Remark

We can compute $\bar{\rho}((S)_{n=1}^N)$, but it is not the signature. It has a higher variance: global minimum paired with global maximum.

Functional Central Limit Theorem: i.i.d.observations

Theorem (Gaussian approximation for empirical processes[9])

Suppose that $X_1, \ldots, X_{N-M+1} \sim X$ are i.i.d.time series. Then,

$$\sqrt{N-M+1}(\hat{F} - F(X)) \to G \quad \text{in distribution}, \tag{13}$$

where G is a Gaussian process with $\mathbb{E}[G_t] = 0$ and covariance

$$(s,t) \mapsto \mathbb{E}[\bar{\rho}_t \bar{\rho}_s] - \mathbb{E}[\bar{\rho}_s]\mathbb{E}[\bar{\rho}_t].$$

In addition, the bootstrap with replacement approximates the limiting distribution, that is $\sqrt{N-M+1}(\hat{F}^*_{N-M+1} - \hat{F}_{N-M+1})$ converges to G.

[9] Frédéric Chazal et al. (2014). "Stochastic Convergence of Persistence Landscapes and Silhouettes". en. In: *Annual Symposium on Computational Geometry - SOCG'14*. Kyoto, Japan, pp. 474–483. (Visited on 03/05/2021).

Functional Central Limit Theorem: i.i.d.observations

Theorem (Gaussian approximation for empirical processes[9])

Suppose that $X_1, \ldots, X_{N-M+1} \sim X$ are i.i.d.time series. Then,

$$\sqrt{N-M+1}(\hat{F} - F(X)) \to G \quad \text{in distribution}, \tag{13}$$

where G is a Gaussian process with $\mathbb{E}[G_t] = 0$ and covariance

$$(s,t) \mapsto \mathbb{E}[\bar{\rho}_t \bar{\rho}_s] - \mathbb{E}[\bar{\rho}_s]\mathbb{E}[\bar{\rho}_t].$$

In addition, the bootstrap with replacement approximates the limiting distribution, that is $\sqrt{N-M+1}(\hat{F}^*_{N-M+1} - \hat{F}_{N-M+1})$ converges to G.

$\implies$ confidence bands and control of type I error in statistical tests!

[9] Frédéric Chazal et al. (2014). "Stochastic Convergence of Persistence Landscapes and Silhouettes". en. In: *Annual Symposium on Computational Geometry - SOCG'14*. Kyoto, Japan, pp. 474–483. (Visited on 03/05/2021).

Dependent observations

There are two levels of dependence in $X_1, X_2, \dots, X_N$

1. Shared elements: $(X_1)_5 = S_5 = (X_4)_2$.
2. Dependence in $(S_n)_n$:
   2.1 γ_{n+1} might not be independent from γ_n.
   2.2 Same for W.

Model

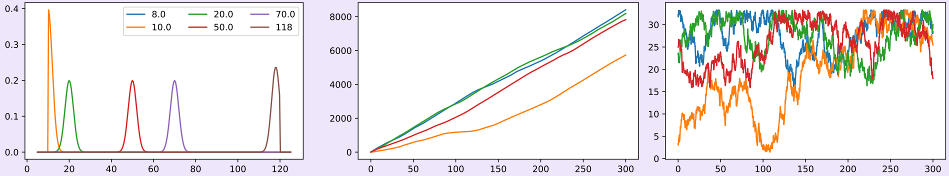
We assume that γ is a Markov Chain, whose first order difference is a Markov Chain, with non-degenerate transitions.

Example

Consider $(V_n)_{n\geq 0}$ be a Markov Chain on $\mathbb{R}$ with

- bounded support $[v_{\min}, v_{\max}]$,
- absolutely continuous transition kernel.

Then, let $\gamma_{n+1} = \gamma_n + \Delta t V_n$.



Functional Central Limit Theorem: dependent observations

Noise

Assume that

1. W is uniformly bounded by $(\max \psi - \min \psi - \epsilon)/2$.
2. There exists $K \in \mathbb{N}$ such that W_n is independent of W_{n+k}, for any $k \geq K$.

Reparametrisation

Fix $M > \frac{2}{v_{\min}}$, where $v_{\min} > 0$ is such that $v_{\min}|t - s| \leq |\gamma(t) - \gamma(s)|$.

Theorem (In progress)

Then, $\sqrt{N - M + 1}(\hat{F}(t) - F(t))$ *converges to a tight, zero–mean Gaussian process* G_d *with covariance*

$$(s, t) \mapsto \lim_{k \to \infty} \sum_{n=1}^{\infty} \operatorname{cov}(\bar{\rho}(X_k)(s), \bar{\rho}(X_n)(t)). \tag{14}$$

Then, if $L(N) \to \infty$ *and* $L(N) = O(N^{1/2-\epsilon})$ *for some* $\epsilon > 0$*, as* $N \to \infty$*,*

$$\sqrt{N - M + 1}(\hat{F}^* - \hat{F}) \to^* G_d(t) \qquad \textit{in probability,}$$

Idea of the proof

For simplicity, assume that the period of ψ is 1.

1. Only the fractional part of $(\gamma_n)_n$ matters: $\psi(\mathrm{frac}(\gamma)) = \psi(\gamma)$.
2. Under the Markov Chain assumptions, $(\mathrm{frac}(\gamma_n))_n$ is β-mixing, with exponential decay rate.
3. By measurability of $x \mapsto \psi(x) + W$, $(S_n)_{n\geq 0}$ is also mixing.
4. Apply functional CLTs for β-mixing data[10] (control the covering number of F_t with respect to $\|\cdot\|_\infty$).

[10] Michael R. Kosorok (2008). *Introduction to Empirical Processes and Semiparametric Inference*. en. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-74977-8 978-0-387-74978-5. DOI: 10.1007/978-0-387-74978-5. URL: http://link.springer.com/10.1007/978-0-387-74978-5 (visited on 10/18/2021); Dragan Radulović (Dec. 1996). "The Bootstrap for Empirical Processes Based on Stationary Observations". In: *Stochastic Processes and their Applications* 65.2, pp. 259–279. ISSN: 0304-4149. DOI: 10.1016/S0304-4149(96)00102-0; Emmanuel Rio (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Springer. DOI: 10.1007/978-3-662-54323-8.

Moving Block Bootstrap

Define windows $(X_n)_{n=1}^{N-M+1}$ as

$$X_n = (S_n, \dots S_{n+M-1}).$$

Let $L = L(N-M+1)$ be a block size

1. $\hat{F} = \frac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(X_n)$.
2. Let $Y_n = (X_n, \dots, X_{n+L})$, for $n = 1, \dots, N-M+1-L$.
3. Let B such that $LB = N+M-1$
4. For $n_b = 1, \dots, n_{\text{bootstrap}}$,
   4.1 Sample $K_1, \dots, K_B \sim \mathcal{U}([1, N-M+1-L])$
   4.2 Obtain $Y_{K_1}, \dots Y_{K_B}$, that is
   $$X_{K_1}, \dots X_{K_1+L}, X_{K_2}, \dots, X_{K_2+L}, X_{K_3}, \dots, X_{K_{B-1}+L}, X_{K_B}, \dots, X_{K_B+L}.$$
   4.3 Calculate $\hat{F}^*_{n_b} = \frac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(X^*_n)$.
5. Calculate some statistics $T(\hat{F}^*_1, \dots, \hat{F}^*_{n_{\text{bootstrap}}}, \hat{F})$.

Hypothesis testing: S_1 vs S_2

Setting

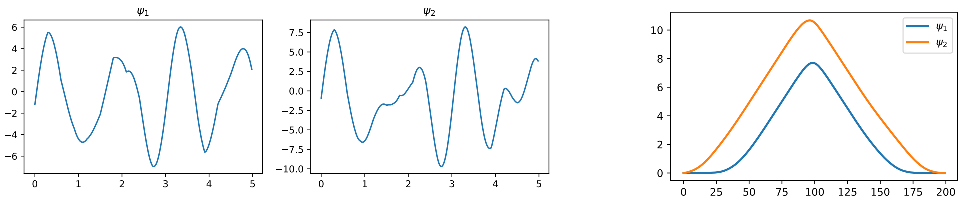
For $k = 1, 2$, we generate $(S_n^k) = \psi_k(\gamma_n) + W_n$.

- $(W_n)_n \sim \mu$ Gaussian, with covariance $(n_1, n_2) \mapsto \exp(-(n_1 - n_2)^2/(2\tau^2))$, with $\tau = 0.008$.
- $\gamma \sim \nu_k$ a Markov Chain of order 3 (acceleration a random walk).
- $V \in [5m/s, 10m/s]$

We have $T = 300s$, and $N = T/dt$, with $dt = 50$.

We test for

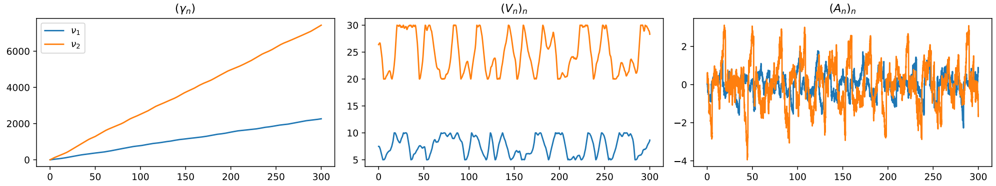
$$H_0 : \ S_1 = S_2 \quad \text{vs} \quad H_1 : S_1 \neq S_2.$$



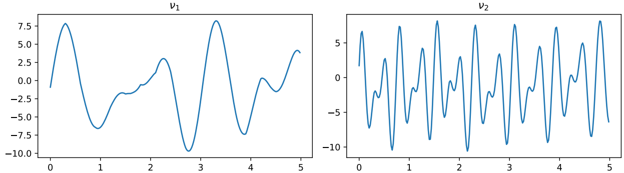
Signatures for $\mu_1 \neq \mu_2$

For $k = 1, 2$, we generate $(S_n^k) = \psi_k(\gamma_n^k) + W_n$.

- $(W_n)_n \sim \mu$ Gaussian, with covariance $(n_1, n_2) \mapsto \exp(-(n_1 - n_2)^2/(2\tau^2))$, with $\tau = 0.008$.
- $\gamma \sim \nu_k$ a Markov Chain of order 3 (acceleration a random walk).
- $V^1 \in [5, 10]$, $V^2 \in [20, 30]$



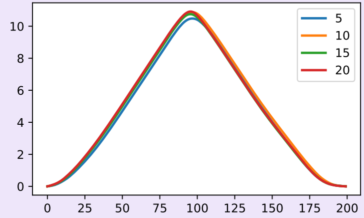
We show samples of X^k for a fixed value of $M = 5s$ and the same ψ below.



Signatures for $\mu_1 \neq \mu_2$

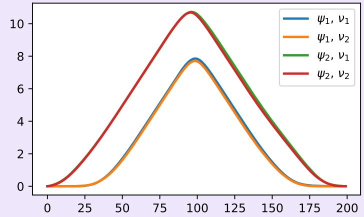
Choice of M

Since we observe a different number of periods in a window of fixed length, the choice of M matters! We inspect the silhouette for different values of M for ν_1.



Test for $\psi^1 = \psi^2$

We fix $M = 10$. The null hypothesis is always rejected - confidence bands too small.



Hypothesis testing: real data

Setting

Scenarios (Segment, direction, Speed), where

- Segment $\in \{A, B\}$,
- Direction $\in \{+, -\}$
- Speed $\in \{10, 30, 50\}$ (km/h).

	Sample size	Description	Example
H_0	12	same scenario and direction, but different speed	(A, $+$, 10) vs (A, $+$, 30)
H_1	66	different scenario or direction	(A, $+$, 10) vs (B, $+$, 30)

Hypothesis testing: real data

Setting

Scenarios (Segment, direction, Speed), where

- Segment $\in \{A, B\}$,
- Direction $\in \{+, -\}$
- Speed $\in \{10, 30, 50\}$ (km/h).

	Sample size	Description	Example
H_0	12	same scenario and direction, but different speed	(A, $+$, 10) vs (A, $+$, 30)
H_1	66	different scenario or direction	(A, $+$, 10) vs (B, $+$, 30)

Results

	Positive	Negative
H_0	1/3	2/3
H_1	1	0

The test is too sensitive: 33% false positives, despite a desired level of 0.5%.

- Attenuation is visible, even with a high sampling rate (125Hz).
- Perturbations of the pattern: signature of traversing a bump different depends on the angular position of the wheel.

Conclusions and future work

Conclusion

We propose topological signatures, F, as invariants of periodic-like processes. We showed that these signatures

1. are invariant to the distribution of reparametrisation, in the noiseless and noisy scenarios,
2. can be estimated from time-series data.

Perspectives

1. Generalize Theorem 1 (convergence of $\bar{\rho}$) to $W \neq 0$.
2. Refine Proposition 1 (continuity with respect to $\nu \sim \gamma$) to different numbers of periods.
3. Study the choice of the window length M, as a function of the number of samples.
4. Application: what it works on?

Conclusions and future work

Conclusion

We propose topological signatures, F, as invariants of periodic-like processes. We showed that these signatures

1. are invariant to the distribution of reparametrisation, in the noiseless and noisy scenarios,
2. can be estimated from time-series data.

Perspectives

1. Generalize Theorem 1 (convergence of $\bar{\rho}$) to $W \neq 0$.
   Difficulty: describe $D(\psi + W)$ based on $D(\psi)$ and W.
2. Refine Proposition 1 (continuity with respect to $\nu \sim \gamma$) to different numbers of periods.
   Difficulty: Describe $D(S)$ based on the knowledge of diagrams on sub-intervals.
3. Study the choice of the window length M, as a function of the number of samples.
4. Application: what it works on?

Adams, Henry and Michael Moy (May 2021). "Topology Applied to Machine Learning: From Global to Local". In: *Frontiers in Artificial Intelligence* 4, p. 668302. ISSN: 2624-8212. DOI: 10.3389/frai.2021.668302.

Adams, Henry et al. (Jan. 2017). "Persistence Images: A Stable Vector Representation of Persistent Homology". en. In: *The Journal of Machine Learning Research* 18.1, pp. 218–252.

Berry, Eric et al. (Apr. 2018). "Functional Summaries of Persistence Diagrams". en. In: *arXiv:1804.01618 [stat]*. arXiv: 1804.01618. URL: http://arxiv.org/abs/1804.01618 (visited on 03/21/2019).

Bonis, Thomas et al. (May 2022). "Topological Phase Estimation Method for Reparameterized Periodic Functions". In: DOI: 10.48550/arXiv.2205.14390.

Bubenik, Peter (Jan. 2015). "Statistical Topological Data Analysis using Persistence Landscapes". en. In: *Journal of Machine Learning Research* 6, pp. 77–102.

Carrière, Mathieu et al. (Mar. 2020). "PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures". In: *arXiv:1904.09378 [cs, math, stat]*. arXiv: 1904.09378 [cs, math, stat].

Chazal, Frédéric et al. (2014). "Stochastic Convergence of Persistence Landscapes and Silhouettes". en. In: *Annual Symposium on Computational Geometry - SOCG'14*. Kyoto, Japan, pp. 474–483. (Visited on 03/05/2021).

Chazal, Frédéric et al. (2016). *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics 2191-8198. Springer, Cham. ISBN: 978-3-319-42543-6.

Kosorok, Michael R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. en. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-74977-8 978-0-387-74978-5. DOI: 10.1007/978-0-387-74978-5. URL: http://link.springer.com/10.1007/978-0-387-74978-5 (visited on 10/18/2021).

Marron, J. S. et al. (Nov. 2015). "Functional Data Analysis of Amplitude and Phase Variation". In: *Statistical Science* 30.4, pp. 468–484. ISSN: 0883-4237. DOI: 10.1214/15-STS524. arXiv: 1512.03216.

Radulović, Dragan (Dec. 1996). "The Bootstrap for Empirical Processes Based on Stationary Observations". In: *Stochastic Processes and their Applications* 65.2, pp. 259–279. ISSN: 0304-4149. DOI: 10.1016/S0304-4149(96)00102-0.

Rio, Emmanuel (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Springer. DOI: 10.1007/978-3-662-54323-8.

Srivastava, A et al. (July 2011). "Shape Analysis of Elastic Curves in Euclidean Spaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.7, pp. 1415–1428. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.184.

Thank you!