Topological signatures of periodic-like signals

FRÉDÉRIC CHAZAL[1,a], BERTRAND MICHEL[2,c] and WOJCIECH REISE[1,b]

[1] *DataShape, Inria Saclay, Institut Mathématique d'Orsay, 307 Rue Michel Magat, 91400 Orsay.,*
[a] *frederic.chazal@inria.fr,* [b] *reisewojciech@gmail.com*
[2] *Nantes Université, Centrale Nantes, Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629 1 Rue de La Noe, 44321 Nantes.,* [c] *bertrand.michel@ec-nantes.fr*

We present a method to construct signatures of periodic-like data. Based on topological considerations, our construction encodes information about the order and values of local extrema. Its main strength is robustness to reparametrisation of the observed signal, so that it depends only on the form of the periodic function. The signature converges as the observation contains increasingly many periods. We show that it can be estimated from the observation of a single time series using bootstrap techniques.

Keywords: Dependent data; functional data; limit theorems; persistent homology; time series

1. Introduction

We consider the problem of constructing a descriptor of a periodic function $\phi : \mathbb{R} \to \mathbb{R}$, based on an observation of a reparameterized and noisy signal. Specifically, we assume that ϕ is 1-periodic and we let $\gamma : [0,T] \to [0,R]$ be an increasing bijection. We consider an observation S of the form

$$S : [0,T] \to \mathbb{R}, \qquad t \mapsto (\phi \circ \gamma)(t) + W(t). \tag{1}$$

where W is a stochastic noise process defined on $[0,1]$. Our aim is to construct a signature $F : S \mapsto F(S)$ which contains information about ϕ while remaining robust to W and to changes in γ. To illustrate this objective more concretely, let's consider the two signals ϕ_0 and ϕ_1 on the top of Figure 1 and the two parametrization functions γ_0 and γ_1 on the left of Figure 1. We would like to define a signature F such that $F(\phi_0 \circ \gamma_0 + W)$ is more similar to $F(\phi_0 \circ \gamma_1 + W)$ than to $F(\phi_1 \circ \gamma_0 + W)$, because the former are both observations of ϕ_0, and despite the fact that $\phi_0 \circ \gamma_0$ and $\phi_0 \circ \gamma_1$ feature three and four periods of ϕ_0 respectively.

Time series or functional observations of the form (1) appear in many applications, where ϕ is somehow characteristic of a population: child growth dynamics (Ramsay and Silverman, 2002), physiological signals (Goldberger et al., 2000), bird migration curves (Su et al., 2014). The reparametrisation γ is the main source of variability in the point-wise evaluations of the signals, as in the 'phase variation' model in Functional data analysis (FDA), see Marron et al. (2015) for a review. The problems typically considered in FDA consist in aligning a population of curves or computing a representative curve, for which methods with guarantees have been proposed (Gasser and Wang, 1997, Khorram, McInnis and Provost, 2019, Tang and Muller, 2008). Underlying most of the models is the assumption that the start and end points ($\gamma(0)$ and $\gamma(T)$ here) are common for all curves.

In applications like magnetic odometry or gait analysis studied in Bonis et al. (2022), Reise (2023) and Bois et al. (2022) respectively, a single observation is composed of several periods of ϕ and the number of periods varies across observations. In Reise (2023), the signal S is the magnetic signal recorded in a moving car and the problem consists in inferring its displacement. The periodic function ϕ models the magnetic signature of the angular position, γ, of that cars' wheel. The problem consists in estimating γ from S. There is little reason for two observations to have the same number of periods,

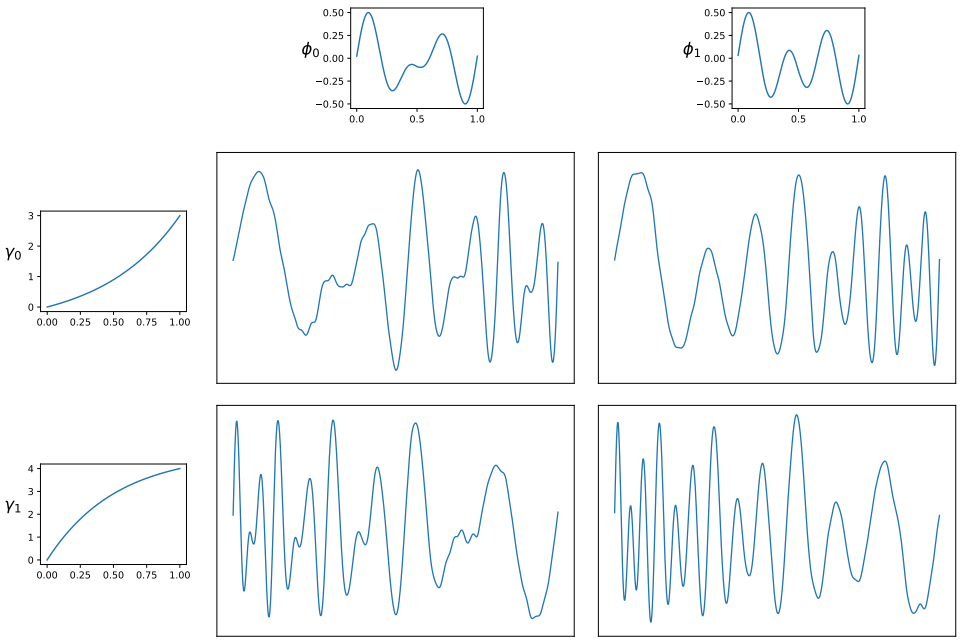

Figure 1. Four signals $S_{i,j} = \phi_j \circ \gamma_i + W$ for $i, j \in \{0, 1\}$, for periodic functions ϕ_0, ϕ_1 and reparametrisations γ_0, γ_1, corrupted by additive noise.

unless the initial angular position of the wheel and the trajectory are exactly the same across those two observations. Therefore, in contrast with FDA, the assumption of common endpoints is not satisfied and the problem changes from describing the whole signal, to that of describing its constituent parts, that is, the periods of ϕ.

Techniques from topological data analysis (TDA) are said to describe the 'shape of data' and have been increasingly used to extract geometric or topological information from observations (Chazal and Michel, 2021). The arguably most popular TDA technique for analyzing a time series consists in computing the homology of the time-delay embedding (TDE) of the time series, in order to verify whether the underlying phenomenon is periodic or not (Perea, 2019). In applications, it has also been used to understand dynamical systems behind climate change (Ghil and Sciamarella, 2023), to identify market crashes (Gidea and Katz, 2018) or to propose biomarkers to detect seizures (Fernández and Mateos, 2022). The TDE of a time series $\mathbf{X} = (X_n)_{n=1}^N$ is a point cloud in $\mathbb{R}^d$, where each point is of the form $(X_n, X_{n+\tau}, \ldots, X_{n+(d-1)\tau})$ for parameters $d, \tau \in \mathbb{N}$. If S is periodic, a simplicial complex constructed on the TDE at the right scale will have a non-trivial homology group in dimension one. In signals with phase variation however, the length of the periodic structure changes and so does the geometry of the TDE. This is corroborated by the fact that the geometry of the delay embedding contains information about the frequencies supporting the signal (Perea, 2019, section 5).

Techniques other than the TDE have been proposed to extract topological information from time series. In Corcoran and Jones (2017), the swarm behavior over time has been described with the zigzag persistent homology of sublevel sets of a density estimator. In Khasawneh and Munch (2016), the authors count revolutions of a machine in an industrial process by counting the number of 'significant' changes in a binary signal, where the significance of a change is defined in terms of persistence of homology generators.

Visual features like local extrema or inflection points (Perng et al., 2000) quantify the 'shape of a curve'. Local extrema and excursion sets are particularly useful, since they are invariant to the reparametrisation of the domain. We propose to use the persistent homology of sublevel sets of the signal to describe this last. This descriptor, the persistence diagram, summarizes the height, order and number of local extrema.

In many statistical applications, it is convenient to map a persistence diagram to a vector or a function, via a functional representation. Numerous functionals (Adams et al., 2017, Carrière et al., 2020) are 'linear in the diagram' and their properties have been well-studied in Divol and Polonik (2019). In our case, it seems natural to renormalize the functionals by the total persistence of the diagram, a proxy for the unknown number of periods. Building on Divol and Polonik (2019) and a recent characterization of the stability of total persistence for Hölder regular processes from Perez (2022), we study the robustness of the signatures we propose.

Guarantees on the estimation of functionals of persistence diagrams, in both asymptotic and non-asymptotic cases, have been provided in Berry et al. (2018), Chazal et al. (2014), under the assumption that the persistence diagrams (or functionals thereof) in the collection are all independent. In a setting motivated by magnetic odometry problem (Bonis et al., 2022), we have a single time series of which we would like to estimate the signature. The natural procedure is to construct a sample by taking contiguous vectors from that observation, what leads to a collection of shorter and dependent observations. We study two reparametrization models and, building on the theory of strong mixing (Dedecker, 2007, Doukhan, 1995), we show that the dependence between observations decreases. When the β-mixing coefficients decrease sufficiently fast, the estimators of the functionals also converge in the dependent setting (Bühlmann, 1995, Kosorok, 2008, Radulović, 1996), not unlike in the independent case (Chazal et al., 2014). So far, estimation of topological signatures from dependent data has been less explored: a concentration inequality for persistent Betti numbers from dependent data is derived in Krebs (2021).

Contributions and outline

In this article, we propose a signature of data of the form (1). The signature captures information about local extrema and excursion sets. It is defined in the language of persistent homology of sublevel sets of a one-dimensional function. Our contribution is an analysis which covers the entire process, from data generation to a bootstrap procedure.

1. By analysing the noiseless case $W = 0$, we show that the proposed signature is a relevant descriptor. First, the topological functional used in the signature converges as the number of periods in a signal grows (Theorem 2.13). Second, by construction, it is invariant to reparametrisation.
2. We discuss the invariance of the signature to parametrization γ in presence of noise. When the the endpoints $\gamma(0)$ and $\gamma(T)$ of the parametrization γ are fixed, an assumption common in FDA, we show that the signature is continuous with respect to the distribution of γ (Theorem 3.4).
3. We consider the problem of estimating the signature from stationary time-series data. When the reparametrisation satisfies Markovian properties, we exploit the periodicity of ϕ to generate a sample to estimate the signature.

In Section 4, we provide a simple numerical illustration of the signature and its invariance properties.

2. Persistence diagrams of sublevel sets and functionals thereof

The signatures we propose are based on excursions sets of stochastic processes. The construction uses the theory of persistent homology, which we describe in the deterministic setting in this section. We

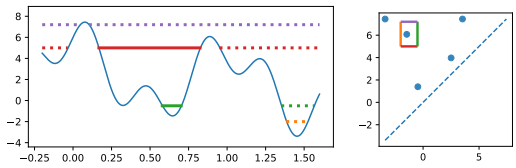

Figure 2. On the left, the graph of a function h and four sublevel sets marked with colors. At each level, there is a different number of connected components. The right panel shows the persistence diagram $D(h)$. The point contained in the interior of the small rectangle corresponds to the connected component marked in a solid line on the left panel.

first introduce a truncated version of persistence to keep boundedness and guarantee continuity of the persistence-based signatures introduced further. Motivated by some additivity property of persistence when studying periodic signals, we consider signatures which are normalized by the total persistence. The objects studied in this section are interesting in themselves. We will therefore adopt a more general setting than the specific model (1).

2.1. Persistence diagrams and total truncated persistence

In this section, we briefly recall the basics of ordinary persistence theory. A more formal presentation of persistent diagrams based on terminated modules is proposed in Section A.1 of the Supplementary Material.

Let $h \in C(\mathbb{X}, \mathbb{R})$ be a continuous function on a compact topological space $\mathbb{X}$. The *α-sublevel set* of h on $\mathbb{X}$ is then defined as: $\mathbb{X}_\alpha = \{x \in \mathbb{X} : h(x) \leq \alpha\}$. Generally speaking, ordinary persistence keeps track of the times of appearance and disappearance of topological features in this sequence. In general, these features can be connected components, loops, cavities, etc. In this work, $\mathbb{X} \subset \mathbb{R}$ so we focus on connected components (0-dimensional features). To fix the ideas, assume that we store the value α_b, called the *birth time*, for which a new connected component appears in $\mathbb{X}_{\alpha_b}$. This connected component eventually gets merged with another one for some value $\alpha_d \geq \alpha_b$, which is stored as well and called the *death time*. Moreover, one says that the component *persists* on the corresponding interval $[\alpha_b, \alpha_d]$. This family of intervals is called the barcode, or *persistence diagram*, of $(\mathbb{X}, h)$, and can be represented as a multiset of points (i.e., point cloud where points are counted with multiplicity) supported on $\mathbb{R}^2$ with coordinates $\{(\alpha_b, \alpha_d)\}$. See Figure 2 for an illustration.

Since the persistence diagram captures information about the number of connected components and how they evolve, it does not depend on the parametrisation of the domain.

Lemma 2.1 (Invariance to reparametrisation). *Consider a continuous function $f : \mathbb{R} \to \mathbb{R}$ and let $\gamma_1, \gamma_2 : [0, T] \to \mathbb{R}$ be two increasing and continuous functions, such that $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(T) = \gamma_2(T)$. Then,*

$$D(f \circ \gamma_1) = D(f \circ \gamma_2).$$

Lemma 2.1 is a consequence of the fact that $\gamma_2 \circ \gamma_1^{-1}$ bijectively maps the connected components of $f \circ \gamma_1$ to those of $f \circ \gamma_2$. A proof is included in Section A.2 of the Supplementary Material.

Persistence diagrams are also stable with respect to the filter function. One distance which is often used to compare diagrams is the bottleneck distance

$$d_b(D_1, D_2) = \inf_{\Gamma} \sup_{x \in D_1 \cup \Delta} \|x - \Gamma(x)\|_\infty,$$

where $\Gamma : D_1 \cup \Delta \to D_2 \cup \Delta$ is a partial bijection between the two diagrams. which allows some points to be matched to the diagonal Δ. With respect to the supremum norm between functions, the persistence diagram is stable in that distance.

Theorem 2.2 (Chazal et al. (2016), Cohen-Steiner, Edelsbrunner and Harer (2007)). *For two functions $f, g : \mathbb{X} \to \mathbb{R}$ with persistence diagrams D_f and D_g respectively,*

$$d_b(D_f, D_g) \leq \|f - g\|_\infty.$$

The *persistence* of $(b, d) \in \mathbb{R}^2$ is $w(b, d) := d - b$. For $p \in \mathbb{N}^+$, the total p-persistence of a persistence diagram D is the sum of p-powers of the lifetimes of points,

$$\text{pers}_p(D) = \left(\sum_{(b,d) \in D} w(b, d)^p \right)^{1/p}.$$

It is similar to total variation for functions on the interval (Plonka and Zheng, 2016). Similarly to total variation, the total p-persistence of an α-Hölder functions is finite for $p > 1/\alpha$, but it is not continuous for functions with regularity strictly less than Lipschitz (Perez, 2022). To keep boundedness and guarantee continuity, we introduce the truncated persistence, $w_\epsilon(b, d) := (d - b - \epsilon)_+$, where $(a)_+ = \max(a, 0)$ denotes the positive par and $\epsilon > 0$ is fixed. For $\varepsilon > 0$, the ϵ-truncated total p-persistence is

$$\text{pers}_{p,\epsilon}(D) = \left(\sum_{(b,d) \in D} w_\epsilon(b, d)^p \right)^{1/p}$$

and $\text{supp}(w_\epsilon) \subset \Delta_\epsilon := \{(b, d) \in \mathbb{R}^2 \mid d - b \geq \epsilon\}$.

We first give two general results on the truncated persistence of continuous functions defined on an interval, before focusing to the case of periodic functions. In the statement of Proposition 2.3, both f and g are determinisitic functions.

Proposition 2.3 (Lower- and upper-bound on ϵ-truncated $p-$persistence). *For continuous functions $f, g : [0, T] \to \mathbb{R}$, and for $p > 0$, $\varepsilon > 0$,*

$$\text{pers}^p_{p,\epsilon}(f + g) \geq \text{pers}^p_{p,\epsilon + A_g}(f),$$

where $A_g = \max g - \min g$. In addition, if f is α-Hölder function with constant Λ and the exponent satisfies $(p - 1)\alpha > 1$, then

$$\text{pers}^p_{p,\epsilon}(f) \leq (A_f - \epsilon)^p_+ \left(1 + pT \left(\frac{2\Lambda}{\epsilon} \right)^{1/\alpha} \right) =: C_{p,\Lambda,\alpha,T}.$$

The result is weak but it is tight. For example, for the lower-bound, if we take f such that $\max f - \min f = 2\|f\|_\infty$ and $g = -\alpha f$, then $f + g = (1 - \alpha) f$ and $\|g\|_\infty = \alpha \|f\|_\infty$, so that $\text{pers}^p_{p,\epsilon}((1 - \alpha) f) = \text{pers}^p_{p,\epsilon + 2\alpha}(f)$.

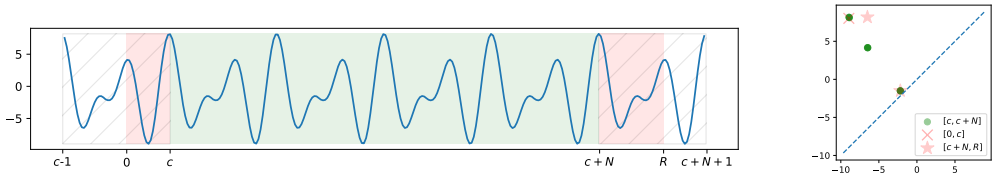

Figure 3. On the left, a graph with of several periods of a periodic function, observed on $[0, R]$ for $R = 5$. The diagram D_1 appears 4 times, and the corresponding parts are marked in green on the graph. The red parts correspond to the remainder D'. The intervals $[c-1, c]$ and $[R, c+N+1]$ included as hatched regions are used to bound the persistence of the remainder. On the left, a graph showing D_1 and D' in green and red respectively.

Proposition 2.4 (Continuity of ϵ-truncated p–persistence). *For any $\epsilon > 0$, the total ϵ-truncated p-persistence $\mathrm{pers}^p_{p,\epsilon} : C([0,T],\mathbb{R}) \to \mathbb{R}$ is continuous. In addition, $\mathrm{pers}^p_{p,\epsilon}$ is Lipschitz over Hölder functions: for any two α-Hölder functions f, g with constant Λ and such that $p - 2 > 1/\alpha$,*

$$\begin{aligned}|\mathrm{pers}^p_{p,\epsilon}(f) - \mathrm{pers}^p_{p,\epsilon}(g)| &\leq p\|f-g\|_\infty \left(\mathrm{pers}^{p-1}_{p-1,\epsilon}(f) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(g)\right) \\ &\leq C_{p-1,\Lambda,\alpha,T}\|f-g\|_\infty.\end{aligned}$$

The proofs of Propositions 2.3 and 2.4, included in Section A of the Supplementary Material, rely on bounds on the maximal number of oscillations of size at least ϵ.

We now focus on the case of periodic function, for which the structure of a persistence diagram can be characterized more finely. Consider $\phi : \mathbb{R} \to \mathbb{R}$ a 1-periodic and continuous function. We denote by $\phi|_A$ the restriction of ϕ to $A \subset \mathbb{R}$ and by $D \sqcup D'$ the union of two multisets. For $n \in \mathbb{N}^*$, nD is the multiset with the same support as D and with multiplicites increased by a factor of n.

Lemma 2.5 (Additivity of diagrams). *Let $R > 1$. There exists $c \in [0,1[$ such that*

$$D(\phi|_{[0,R]}) = \left(\bigsqcup_{k=1}^{\lfloor R-1 \rfloor} D(\phi|_{[c+k-1,c+k]})\right) \sqcup D', \tag{2}$$

for some persistence diagram D'. In addition, $D_1 \coloneqq D(\phi|_{[c,c+1]}) = D(\phi|_{[c+k-1,c+k]})$ for any k and $\mathrm{pers}_{p,\epsilon}(D') \leq 2\mathrm{pers}_{p,\epsilon}(D_1)$ for any $\epsilon \geq 0$.

We defer the full proof to Section A of the Supplementary Material, but we provide an illustration of the core idea in Figure 3. By choosing $c \in [0,1]$ to be a global maximum of ϕ and defining "the period" to be $\phi|_{[c,c+1]}$, we can decompose the diagram of $\phi|_{[0,R]}$ as a sum of diagrams of individual periods. These diagrams are all equal and denoted D_1, except for the first and last periods which are not complete and constitute the residual part D'. Thus, we obtain (2). The total persistence of a diagram of an incomplete period is not greater than the persistence of a complete period, so $\mathrm{pers}_{p,\epsilon}(D') \leq 2\mathrm{pers}_{p,\epsilon}(D_1)$.

In the case of persistence homology of sublevel sets of periodic functions, most points in the persistence diagram will have multiplicity greater than one, reflecting the number of observed periods, as stated by Lemma 2.5. When the function is corrupted by additive noise, the points will no longer be superposed. The additive structure combined with robustness in the form of to the bottleneck stability motivates us to introduce normalized versions of persistence-based signatures.

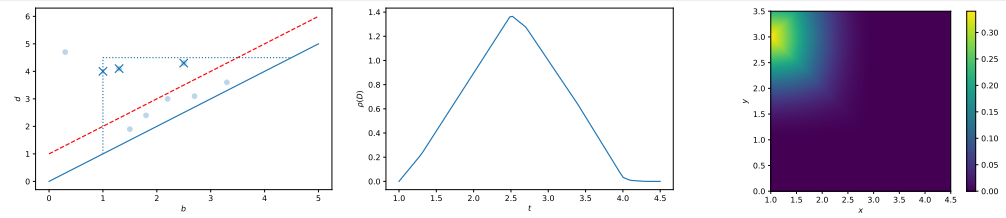

Figure 4. On the right, a persistence diagram with $\epsilon = 1$, $L = 1$ and $U = 4.5$ marked in dashed and dotted lines. In the center and on the right, two functionals $\bar{\rho}$ evaluated on that diagram, with k^s and $k^{pi,t}$ ($\sigma = 1, r = 1.1$) respectively, both weighted by the truncated persistence w_ϵ with $p = 2$.

2.2. Normalized functionals of persistence

The space of persistence diagrams is not a vector space and is ill-suited for statistical purposes. It is common to map diagrams to a functional Banach space. Many such mappings have been proposed (Adams et al., 2017, Bubenik, 2015, Carrière et al., 2020, Chung and Lawson, 2022) and their properties are described extensively. For the purpose of creating signatures of periodic functions, functionals normalized by the total persistence are of particular interest. As it is usually the case in the TDA literature, we present a general set of assumptions and we show examples of functionals from the literature (or of their adaptation) which fit within the prescribed framework.

Consider $\mathbb{U}$ a Euclidean space and let $\mathcal{H}$ be a Banach space of functions $\mathbb{U} \to \mathbb{R}$. Let $k : \mathbb{R}^2 \to \mathcal{H}$ be a map, which to a point (b, d) in the plane associates a function $k(b, d) : \mathbb{U} \to \mathbb{R}$. We give two examples, the persistence silhouette kernel $\Lambda_{(b,d)}$ from Bubenik (2015) and the persistence image kernel k^{pi} from Adams et al. (2017),

$$\Lambda_{(b,d)} : \mathbb{R} \to \mathbb{R} \qquad k^{pi}(b,d) : \mathbb{R}^2 \to \mathbb{R}$$
$$u \mapsto \left(\tfrac{d-b}{2} - |u - \tfrac{b+d}{2}|\right)_+, \qquad (x,y) \mapsto \tfrac{1}{2\pi\sigma^2} \exp\left(-\tfrac{(b-x)^2+(d-y)^2}{2\sigma^2}\right).$$

Definition 2.6. For a persistence diagram D with $\mathrm{pers}_{p,\epsilon}(D) > 0$, the linear and the normalized functionals are defined as

$$\rho(D) \coloneqq \sum_{x \in D} w_\epsilon(x)^p k(x), \qquad \bar{\rho}(D) \coloneqq \frac{\rho(D)}{\sum_{x \in D} w_\epsilon(x)^p}. \tag{3}$$

Otherwise, when $\mathrm{pers}_{p,\epsilon}(D) = 0$, we set $\rho(D) = 0 = \bar{\rho}(D)$.

Note that $\rho(D)$ and $\bar{\rho}(D)$ are by construction elements of $\mathcal{H}$, that is functions from $\mathbb{U}$ to $\mathbb{R}$. In this work, we are more specifically interested in diagrams of sublevel sets of functions defined on a compact interval. In this case we will abuse notation and write $\rho(f) \coloneqq \rho(D(f))$.

Remark 2.7. We note a few differences with Persistence Curves introduced in Chung and Lawson (2022). In that article, the operator aggregating $\{k(x)\}_{x \in D}$ is different from the sum used here. In addition, for normalized functionals, the authors only consider kernels of the form $k(b, d)(u) = c 1_{[b,d]}(u)$, for some $c > 0$, which are piecewise constant, so not continuous.

We will assume that k satisfies the following:

1. $k(x)$ has a uniformly bounded support, for all $x \in \mathbb{R}^2$

$$\exists K \subset \mathbb{U} \text{ compact}, k(x)|_{\mathbb{U}\setminus K} \equiv 0, \qquad \forall x. \tag{4}$$

2. $k(x)$ is Lipschitz, uniformly over $x \in \mathbb{R}^2$

$$\exists L > 0,\ |k(x)(u) - k(x)(s)| \leq L d(u,s), \qquad \forall x \in \mathbb{R}^2, \forall u, s \in \mathbb{U}. \tag{5}$$

3. $x \mapsto k(x)$ is Lipschitz

$$\exists L_k > 0, \|k(x) - k(x')\|_{\mathcal{H}} \leq L_k \|x - x'\|_\infty, \qquad \forall\, x, x' \in \mathbb{R}^2. \tag{6}$$

4. $k(x)$ is uniformly-bounded on the diagonal

$$\exists C \leq 0,\ \|k|_\Delta\|_{\mathcal{H}} \leq C. \tag{7}$$

Assumptions (5-7) are standard in the literature. Assumptions (6) and (7) guarantee that $\|k(x)\|_{\mathcal{H}}$ is uniformly bounded on any compact subset of $\mathbb{R}^2$. While a condition on $\|k(x)\|_{\mathcal{H}}$ could be imposed for a different x, an assumption for $x \in \Delta$ is natural because points in Δ correspond to generators in the module which exist punctually and do not persist. However, many kernels do not satisfy (4), and it is often obtained as a consequence of an assumption on the statistical model (Berry et al., 2018, Chazal et al., 2014). To adapt a kernel to this assumption, we can precompose it with a projection. Specifically, let $Q < U \in \mathbb{R}$ and consider $\pi_{Q,U} : \Delta_{\geq 0} \to \Delta_{\geq 0}$ the operator which maps points above the diagonal, onto the upper triangle with corner at (Q, U)

$$\begin{aligned} \pi_{Q,U} :\ & \Delta_{\geq 0} \to \Delta_{\geq 0} \\ & (b,d) \mapsto (b,d) + (1,-1)\min(\max(d-U, Q-b, 0), \tfrac{d-b}{2}). \end{aligned} \tag{8}$$

Examples of such functionals are given below for the Persistence Silhouette and the Persistence Image, and sample realisations are shown in Figure 4. The calculations of the Lipschitz constants are carried out in Section B.1 of the Supplementary Material.

Example 2.8 (Persistence Silhouette). We set $k^s(b,d)(u) = \Lambda_{(\pi_{Q,U}(b,d))}(u)$, so that $\operatorname{supp}(k^s(b,d)) \subset \operatorname{supp}(\Lambda^s(Q,U)) = [Q,U]$. Since $t \mapsto k^s(b,d)$ is piecewise linear with slopes 0, 1 and -1, $L = 2 = L_k$. The kernel is zero on the diagonal, so $C = 0$ is enough to satisfy (7).

Example 2.9 (Persistence Image). In order for k^{pi} to have bounded support and remain Lipschitz, we propose to multiply by it the distance to a square of size 2σ to (b,d), namely, for some $r > 1$, set

$$k^{pi,r}(b,d)(x,y) = \left(2 - \frac{\|\pi_{Q,U}(b,d) - (x,y)\|_\infty}{\sigma}\right)_+^r k^{pi}(\pi_{Q,U}(b,d))(x,y).$$

We obtain the original persistence image kernel when we set $r = 0$ and $Q = \infty, U = \infty$. Note that a simple truncation of k^{pi} is not enough, as the kernel would not be continuous at the truncation interface. The function $(x,y) \mapsto \exp(-(x^2+y^2))$ is $(4/e)$-Lipschitz and $(x,y) \mapsto (2 - \|(b,d) - (x,y)\|_\infty/\sigma)_+^r$ is $(r2^r/\sigma)$-Lipschitz, for the Minkowski distance. Hence, $L_{k^{pi,r}} = 2^{r-1}(r+2)/\pi\sigma^3$ and $L = 2^{r+1}/\pi e\sigma^3$.

Continuity of functionals has been studied, notably in Divol and Polonik (2019) and Chung and Lawson (2022). In the first, it was fully characterized, but only for linear functionals. In the latter, functionals were considered under the L^1 metric. Due to the nature of the statistical results in Section 3.3, we are particularly interested in $\|\cdot\|_\infty$, so we repeat the proof of Divol and Polonik (2019, Theorem 3) for linear functionals ρ and we derive results for normalized functionals $\bar{\rho}$. The proof of Proposition 2.10 is in Section B.2 of the Supplementary Material.

Proposition 2.10 (Stability). *Suppose that the persistence of any point in D_1 and D_2 is bounded by a uniform constant U and that k satisfies (4), (6) and (7). Then,*

$$\|\rho(D_1)-\rho(D_2)\|_{\mathcal{H}} \leq \left(L_k \mathrm{pers}^{p}_{p,\epsilon}(D_1) + p(L_kU+C)\sum_{k=1,2}\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_k)\right) d_B(D_1,D_2), \tag{9}$$

$$\|\bar{\rho}(D_1)-\bar{\rho}(D_2)\|_{\mathcal{H}} \leq \left(L_k + 2p(L_kU+C)\frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1)+\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2)}{\mathrm{pers}^{p}_{p,\epsilon}(D_1)}\right) d_B(D_1,D_2). \tag{10}$$

Remark 2.11. The result we give for ρ is a special case of Divol and Polonik (2019, Theorem 3). To see this, notice that using the notations of that article, $Lip(\phi) = L_k$, $A = p$, and $\alpha = p$, where 'p' is from our work. In their article, $p = \infty$ and $a = 1$. In particular, we see exactly that $\sup_x \|k(x)\|_{\mathcal{H}} \leq L_kU + C$.

Corollary 2.12. *Let $f \in C([0,T],\mathbb{R})$ such that $A_f > \epsilon$. Then, the linear and normalized functionals $\rho(D(f))$ and $\bar{\rho}(D(f))$ in (3) are well-defined. In addition, $h \mapsto \bar{\rho}(D(h))$ is continuous at f for $\|\cdot\|_\infty$.*

Proof of Corollary 2.12. Since f is continuous on a compact domain, it is also uniformly continuous and bounded. Let $\delta > 0$ be such that $|f(t)-f(s)| < \epsilon$, whenever $|s-t| < \delta$. By the reasoning of the proof of (Cohen-Steiner et al., 2010, Persistence Cycle Lemma), $|\omega^{-1}(]\epsilon,\infty[) \cap D(f)| \leq 1 + T/\delta$. Let $M_f = \max(f)$, $m_f = \min(f)$. Then,

$$\begin{aligned}\|\rho(D(f))\|_{\mathcal{H}} &\leq \sum_{(b,d)\in D(f)} w_\epsilon(d-b)^p \|k(b,d)\|_{\mathcal{H}} \\ &\leq (\tfrac{T}{\delta}+1)\cdot w_\epsilon (M_f - m_f)^p \max_{(b,d)\in D(f)\cap\Delta^+_\epsilon} \|k(b,d)\|_{\mathcal{H}}.\end{aligned}$$

As stated above, the number of points is bounded from above, and so is the total persistence, thus showing that $\rho(D(f))$ is well-defined. For the normalized functional,

$$\|\bar{\rho}(D(f))\|_{\mathcal{H}} \leq (\tfrac{T}{\delta}+1) \max_{(b,d)\in D(f)\cap\Delta_{\geq\epsilon}} \|k(b,d)\|_{\mathcal{H}} \leq (\tfrac{T}{\delta}+1) \max_{(b,d)\in D(f)\cap\Delta_{\geq\epsilon}} \|k(b,d)\|_{\mathcal{H}}$$

To show the continuity, let $z > 0$ and consider $h \in C([0,T],\mathbb{R})$ such that $\|f-h\|_\infty < r_f$, where

$$r_f := \min\left(\frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(f)}{4pM_f(A_f+\epsilon/2)^{p-1}}, \frac{z}{L_k+4p(2L_kA_f)\mathrm{pers}^{p-1}_{p-1,\epsilon}(f)/\mathrm{pers}^{p}_{p,\epsilon}(f)}, \frac{A_f-\epsilon}{2}\right),$$

and $M_f = |D(f)\cap\Delta_{\geq\epsilon/2}|$. By continuity of truncated persistence, and with the modulus of continuity from the proof of Proposition 2.4, we have $\mathrm{pers}^{p-1}_{p-1,\epsilon}(h) \leq 2\mathrm{pers}^{p-1}_{p-1,\epsilon}(f)$. We observe that $d-b \leq A_f$ for $(b,d) \in D(f)$ and that $d-b \leq A_h \leq A_f + 2\|f-h\|_\infty \leq 2A_f$. Applying Proposition 2.10 to $D_1 = D(f)$, $D_2 = D(h)$ with $U = 2A_f$, we obtain

$$\|\bar{\rho}(D(f))-\bar{\rho}(D(h))\|_{\mathcal{H}} \leq \left(L_k + 4p(2A_fL_k+C)\frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(f)}{\mathrm{pers}^{p}_{p,\epsilon}(f)}\right) d_B(D(f),D(f)) \leq z.$$

□

2.3. Normalized functionals of persistence for periodic functions

Let us now go back to the persistence diagram of a 1-periodic and continuous function $\phi : \mathbb{R} \to \mathbb{R}$. As a consequence of the additivity of diagrams (Lemma 2.5), we find that the normalized functionals are consistent. Indeed, the functional converges because the contribution of the boundary effects in the normalized functional $\bar{\rho}(\phi|_{[0,R]})$ decreases as the number of observed periods increases to infinity. This justifies calling the limit the "signature of ϕ".

Theorem 2.13 (Consistency). *Assume that k satisfies (6) and (7). Then, there exists $c \in [0,1[$ such that as $R \to \infty$,*

$$\bar{\rho}(D(\phi|_{[0,R]})) \xrightarrow{\|\cdot\|_{\mathcal{H}}} \bar{\rho}(D(\phi|_{[c,c+1]})).$$

Proof. Let $D_1 = D(\phi|_{[c,c+1]})$, D' be given by Lemma 2.5 and let $D_R = D(\phi|_{[0,R]})$. In addition, we will write $\rho(D) = \rho_{k,\epsilon,p}(D) = \sum_{x\in D} w_\epsilon(x)^p k(x)$ for the linear (non-normalized) version of the functional $\bar{\rho}$ from (3). Then,

$$\left\| \frac{\rho((R-1)D_1)+\rho(D')}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \frac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)} \right\|_{\mathcal{H}} \leq \left\| \frac{\rho(D')}{\mathrm{pers}^p_{p,\epsilon}(D_R)} \right\|_{\mathcal{H}} + \left\| \frac{\rho((R-1)D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \frac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)} \right\|_{\mathcal{H}},$$

and

$$\begin{aligned}\left\| \frac{\rho((R-1)D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \frac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)} \right\|_{\mathcal{H}} &\leq \left\| \frac{\mathrm{pers}^p_{p,\epsilon}(D_1)\rho((R-1)D_1) - (\mathrm{pers}^p_{p,\epsilon}((R-1)D_1)+\mathrm{pers}^p_{p,\epsilon}(D'))\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)} \right\|_{\mathcal{H}} \\ &\leq \frac{\|\mathrm{pers}^p_{p,\epsilon}(D')\rho(D_1)\|_{\mathcal{H}}}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)},\end{aligned}$$

where we have used that for any $N \in \mathbb{N}$,

$$\mathrm{pers}^p_{p,\epsilon}(ND_1)\rho(D_1) = N\mathrm{pers}^p_{p,\epsilon}(D_1)\rho(D_1) = \mathrm{pers}^p_{p,\epsilon}(D_1)\rho(ND_1).$$

Now, we observe that $\mathrm{pers}^p_{p,\epsilon}(D_R) = \mathrm{pers}^p_{p,\epsilon}((R-1)D_1) + \mathrm{pers}^p_{p,\epsilon}(D') \geq (R-1)\mathrm{pers}^p_{p,\epsilon}(D_1)$ and $\mathrm{pers}^p_{p,\epsilon}(D') \leq 2\mathrm{pers}^p_{p,\epsilon}(D_1)$ to obtain that

$$\|\bar{\rho}(D(\phi|_{[0,R]})) - \bar{\rho}(D(\phi|_{[c,c+1]}))\|_{\mathcal{H}} \leq \frac{\|\rho(D')\|_{\mathcal{H}}}{\mathrm{pers}^p_{p,\epsilon}(D_R)} + \frac{\|\mathrm{pers}^p_{p,\epsilon}(D')\rho(D_1)\|_{\mathcal{H}}}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)} \leq \frac{\|\rho(D')\|_{\mathcal{H}}+2\|\rho(D_1)\|_{\mathcal{H}}}{(R-1)\mathrm{pers}^p_{p,\epsilon}(D_1)} \tag{11}$$

Using the Minkowski inequality,

$$\|\rho(D')\|_{\mathcal{H}} = \|\sum_{x\in D'} w_\epsilon(x)^p k(x)\|_{\mathcal{H}} \leq \sum_{x\in D'} |w_\epsilon(x)^p| \max_{x\in D'} \|k(x)\|_{\mathcal{H}} \leq \mathrm{pers}^p_{p,\epsilon}(D') \max_{x\in D'} \|k(x)\|_{\mathcal{H}}.$$

Because k is L_k-Lipschitz by (6), for any $x \in D'$, we have $\|k(x)\|_{\mathcal{H}} \leq L_k\|x - \pi(x)\|_\infty + \|k(\pi(x))\|_\infty$, where $\pi(b,d) = ((b+d)/2, (b+d)/2)$. Using (7) on one hand, and the fact that the distance of any point in the diagram to Δ is bounded by A_ϕ, we obtain $\|k(x)\|_{\mathcal{H}} \leq L_k A_\phi/2 + C$. A similar bound holds for $\|\rho(D_1)\|_{\mathcal{H}}$. Going back to (11), we have that

$$\begin{aligned}\|\bar{\rho}(D(\phi|_{[0,R]})) - \bar{\rho}(D(\phi|_{[c,c+1]}))\|_{\mathcal{H}} &\leq \frac{(2|\mathrm{pers}^p_{p,\epsilon}(D_1)| + |\mathrm{pers}^p_{p,\epsilon}(D_1)|)\max_{x\in D'}\|k(x)\|_{\mathcal{H}}}{(R-1)\mathrm{pers}^p_{p,\epsilon}(D_1)} \\ &\leq \frac{4(C + L_k A_\phi)}{R-1},\end{aligned}$$

what converges uniformly to 0 as R tends to infinity. □

We conclude the section with another stability bound of normalized functionals under reparametrization and (deterministic) additive noise.

Proposition 2.14. *Let $(\gamma_k : [0,T] \to [0,R_k])_{k=1,2}$ be two fixed reparametrisations, for $R_k > 2$ and let $g_1, g_2 \in C([0,T],\mathbb{R})$ be two α-Hölder functions with constant Λ_g and with sup-norm bounded as follows $\|g_k\|_\infty < A_\phi/2$. Then,*

$$\|\bar{\rho}(\phi \circ \gamma_1 + g_1) - \bar{\rho}(\phi \circ \gamma_2 + g_2)\|_{\mathcal{H}} \leq L_k \left(\tfrac{4A_\phi}{\min(R_1,R_2)-2} + P(\max(\|g_1\|_\infty, \|g_2\|_\infty)) \right),$$

where $P(x) = O(x)$.

The proof as well as the expression of $P(x)$ can be found in Section B.3 of the Supplementary Material. The Proposition is a straightforward consequence of Proposition 2.10 and Theorem 2.13. Note that the right–hand side is strictly positive, even in the noiseless case $g = 0$. It is not surprising, because the bounds we use are very crude: we remove the noise and we compare the respective signatures to the limit object $\bar{\rho}(\phi)$. In the next section we will provide alternative stability bounds for signatures under reparametrization and additive noise in a stochastic setting.

3. Signatures of periodic signals with phase variation

In this Section we define what we call the signature of a reparametrized periodic function, in a probabilistic model. Next we provide robustness properties of the signature and statistical guarantees for its estimation.

3.1. Generative model and persistence-based signature

As the signature will be defined in term of an expectation, we first need to specify the probabilistic model associated to our main model (1).

- We fix $\phi : \mathbb{R} \to \mathbb{R}$ to be a continuous and 1-periodic function.
- We consider reparametrisations which have a lower-bound on the modulus of continuity. Specifically, let $\Gamma_{v_{\min}} := \{\gamma \in C([0,T],\mathbb{R}) \mid \gamma(s) - \gamma(t) \geq v_{\min}(s-t), \text{ for all } s \geq t\}$ be the space of reparametrisations, equipped with the Borel σ-algebra $\mathcal{B}(\|\cdot\|_\infty)$ and let μ be a probability measure on that space.
- We consider $C_{A_\phi-(\epsilon+q)}([0,T],\mathbb{R}) := \{W \in C([0,T],\mathbb{R}) \mid A_W \leq A_\phi - (\epsilon+q)\}$ the closed subspace of continuous functions with amplitude bounded by $A_\phi - (\epsilon+q)$ for some $q > 0$ and let ν be a probability measure on $(C_{A_\phi-(\epsilon+q)}([0,T],\mathbb{R}), \mathcal{B}(\|\cdot\|_\infty))$.

Under theses assumptions, we consider the stochastic process

$$S: \qquad t \mapsto \phi(\gamma(t)) + W(t), \qquad \text{where } \gamma \sim \mu \text{ and } W \sim \nu. \tag{12}$$

Here, γ and W are independent. Moreover the process γ is not assumed to be observed. Indeed this is not necessary to observe γ to define our persistence-based signature, which is robust to the temporal changes.

We are now in position to introduce our persistence-based signature built on the signal S. Starting from a kernel k, we define $\bar{\rho}(S)$ as in Equation (3). For a realization $S(w)$ of S and $u \in \mathbb{U}$, and we can calculate $\bar{\rho}(S(w))(u) \in \mathbb{R}$. Since $A_{S(w)} \geq A_\phi - A_W \geq \epsilon + q$, by Corollary 2.12, $\bar{\rho}(S)(u)$ is a bounded real-valued random variable. We then define the signature of S point-wise as

$$F(S)(u) \coloneqq \mathbb{E}[\bar{\rho}(S)(u)] = \mathbb{E}_{\gamma\sim\mu,\, W\sim\nu}[\bar{\rho}(\phi\circ\gamma + W)(u)], \tag{13}$$

where the expectation is taken with respect to the law of the process.

Remark 3.1. Note that T is fixed in our setting, so that the 'time' of the observation is the same, but $\gamma(T) - \gamma(0)$, the number of periods of ϕ in S may vary. This is in contrast with typical assumptions for functions with phase variation, where $\gamma(T) - \gamma(0)$ is shared between observations.

In Section C of the Supplementary Material, we show that $\bar{\rho}(S) \in C(\mathbb{U}, \mathbb{R})$ is also a random process as stated formally in Proposition 3.2.

Proposition 3.2. *Under Model (12), if $\bar{\rho} : C_{\epsilon+q}([0,T],\mathbb{R}) \to C(\mathbb{U},\mathbb{R})$ is continuous and $C(\mathbb{U},\mathbb{R})$ is $\|\cdot\|_\infty$-separable, then $\bar{\rho}(S)$ is $(C(\mathbb{U},\mathbb{R}), \|\cdot\|_\infty)$-measurable.*

3.2. Robustness of the signature to reparametrisation

We now investigate the robustness properties of the signature F, in both the noiseless and the additive noise settings.

3.2.1. Robustness in the noiseless setting

We first consider the noiseless case $W = 0$. Lemma 2.1 implies that the functional depends only on the number of periods. As a consequence, the signature $F(S)$ is also robust to the distribution of reparametrisations, but only when the distributions of endpoints are equal.

Let $\delta_t : \gamma \mapsto \gamma_t$ be the evaluation map defined on $\Gamma_{v_{\min}}$. We introduce $(\delta_{0,T})_\star\mu \coloneqq \mu \circ (\delta_0, \delta_T)^{-1}$ the push-forward measure on $\mathbb{R}^2$, which characterizes the distribution of $(\gamma(0), \gamma(T))$.

Proposition 3.3. *Consider $\mu_1, \sim \mu_2$ two probability measures on $\Gamma_{v_{\min}}$. If we let $\gamma_k \sim \mu_k$ such that $(\delta_{0,T})_\star\mu_1 \overset{\mathcal{L}}{=} (\delta_{0,T})_\star\mu_2$, then*

$$F(\phi\circ\gamma_1) = F(\phi\circ\gamma_2). \tag{14}$$

Proof. We first show that we can condition the measure μ_k on $(\gamma_k(0), \gamma_k(T)) = x$. The space of continuous functions $C([0,T],\mathbb{R})$ is Polish, and so is $\Gamma_{v_{\min}}$ as a closed subspace. By Bogachev (2007, Corollary 10.4.6), there are regular conditional measures $((\mu_1)_x(d\gamma))_{x\in\mathbb{R}^2}$ and $((\mu_2)_x(d\gamma))_{x\in\mathbb{R}^2}$. Lemma 2.1 implies that $\gamma \mapsto \bar{\rho}(\phi\circ\gamma)(u)$ is constant on $\delta_{0,T}^{-1}(x)$, for any $x = (s,r) \in \mathbb{R}^2$. For any $u \in \mathbb{U}$, using the regular conditional measure property Bogachev (2007, Definition 10.4.1),

$$\begin{aligned} F(\phi\circ\gamma_1)(u) &= \int_\Gamma \bar{\rho}(\phi\circ\gamma)(u)\mu_1(d\gamma) \\ &= \int_{\mathbb{R}^2}\int_{\delta_{0,T}^{-1}(x)} \bar{\rho}(\phi\circ\gamma)(u)(\mu_1)_x(d\gamma)(\delta_{0,T})_\star\mu_1(dx) \end{aligned}$$

$$= \int_{\mathbb{R}^2} \int_{\delta_{0,T}^{-1}(x)} \bar{\rho}(\phi \circ \gamma)(u)(\mu_2)_x(d\gamma)(\delta_{0,T})_\star \mu_2(dx)$$
$$= F(\phi \circ \gamma_2)(u).$$

□

While it is disappointing to require equality of the marginals $(\delta_{0,T})_\star \mu_1$ and $(\delta_{0,T})_\star \mu_2$ in Proposition 3.3, removing this assumption poses a difficulty which we now discuss. In short, the main problem lies in obtaining a fine control on the persistence diagram when 'cutting' a domain, $[0, R]$, into $[0, R_1]$ and $[R_1, R_2]$, for any $0 < R_1 < R_2$. Specifically, we need to consider the difference between $D(\phi|_{[0,R_2]})$ and $D(\phi|_{[0,R_1]}) \cup D(\phi|_{[R_1,R_2]})$, and it is not zero unless $R_1 \in \operatorname{argmax}\phi$. When R_1 is a global maximum of ϕ, we can reason as in the proof of Lemma 2.5. However, this is far from the general situation, in which case the cut at R_1 might induce some spurious points in the diagram, as we now illustrate and discuss briefly.

Assume for simplicity that γ_1 and γ_2 are fixed with $R_k := \gamma_k(T)$. Without loss of generality, assume that $R_1 < R_2$. and let $T_1 := \gamma_2^{-1}(R_1)$. Lemma 2.1 implies that $\bar{\rho}(\phi \circ \gamma_1) = \bar{\rho}((\phi \circ \gamma_2)|_{[0,T_1]})$. If we let $D_1 = D((\phi \circ \gamma_2)|_{[0,T_1]})$ and $D_2 = D((\phi \circ \gamma_2)|_{[T_1,T]})$, we obtain that

$$\|\bar{\rho}(\phi \circ \gamma_1) - \bar{\rho}(\phi \circ \gamma_2)\|_{\mathcal{H}} \leq \|\bar{\rho}(D_1) - \bar{\rho}(D_1 \sqcup D_2)\|_{\mathcal{H}} + \|\bar{\rho}(D_1 \sqcup D_2) - \bar{\rho}(\phi \circ \gamma_2)\|_{\mathcal{H}}. \tag{15}$$

We can conveniently analyze the first term of (15) by observing that a normalized functional of a union of diagrams is a weighted average of the normalized functionals of the individual diagrams

$$\bar{\rho}(D_1 \sqcup D_2) = \bar{\rho}(D_1) \frac{\operatorname{pers}_{p,\epsilon}^p(D_1)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} + \bar{\rho}(D_2) \frac{\operatorname{pers}_{p,\epsilon}^p(D_2)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)},$$

so that

$$\|\bar{\rho}(D_1) - \bar{\rho}(D_1 \sqcup D_2)\|_{\mathcal{H}} = \|\bar{\rho}(D_1)\left(\frac{\operatorname{pers}_{p,\epsilon}^p(D_1)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} - 1\right) + \bar{\rho}(D_2)\frac{\operatorname{pers}_{p,\epsilon}^p(D_2)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)}\|_{\mathcal{H}}$$
$$= \|\bar{\rho}(D_1) - \bar{\rho}(D_2)\|_{\mathcal{H}} \frac{\operatorname{pers}_{p,\epsilon}^p(D_2)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)}$$
$$\leq (L_k A_\phi + C) \frac{\operatorname{pers}_{p,\epsilon}^p(D_2)}{\operatorname{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)}.$$

We claim that if ϕ is regular enough and $R_2 - R_1$ is small, then so is $\operatorname{pers}_{p,\epsilon}^p(D_2)$. The first term in (15) is thus not too problematic. On the contrary, controlling the second term in (15) is harder. Thanks to Proposition 2.10, it can be can be upper bounded by $d_B(D_1 \sqcup D_2, D(\phi \circ \gamma_2))$, which, unless R_1 is a global maximum of ϕ, is positive. In conclusion, we see that to get a general stability from (15), we are lacking a tight control of of $d_B(D_1 \sqcup D_2, D(\phi \circ \gamma_2))$.

3.2.2. Robustness in the additive noise setting

Let us go back to the model with noise, introduced in (12). Theorem 3.4 expresses the stability of the signature for different distributions of γ with endpoints $\gamma(0) = 0$, $\gamma(T) = R$ fixed across all reparametrisations. The aim here is to compare the impact of noise on the signature for two different parameterizations. Let $0 < T, Tv_{\min} < R$ and consider

$$\Gamma_{T,R,v_{\min}} := \{\gamma \in C([0,T],[0,R]) \mid \gamma(0) = 0, \gamma(T) = R,\ 0 \leq v_{\min}(t-s) \leq \gamma(t) - \gamma(s), \forall s \leq t\}.$$

The set $\Gamma_{T,R,v_{\min}}$ is convex. It is also included in $C([0,T],\mathbb{R})$, so it can be naturally endowed with the sup-norm, for which it is a closed, complete and separable space. In particular, it is a Radon space, so that all measures on $(\Gamma_{T,R,v_{\min}}, \mathcal{B}(\Gamma_{T,R,v_{\min}}))$ are inner–regular and locally-finite. Hence, we can equip the space of probability measures on $(\Gamma_{T,R,v_{\min}}, \mathcal{B}(\Gamma_{T,R,v_{\min}}))$ with the Wasserstein distance $W_{1,\|\cdot\|_\infty}$ (Panaretos and Zemel, 2020).

Concerning the noise W, in addition to the bound on A_W introduced in (12), we impose a path-wise regularity condition: for some $0 < r_1 < r_2$,

$$\text{there exists } K = K_{r_2,r_1}, \text{ such that } \mathbb{E}[|W_t - W_s|^{r_2}] \leq K_{r_2,r_1}|t-s|^{1+r_1}, \text{ for all } s,t \in [0,T]. \tag{16}$$

Theorem 3.4 (Stability). *Let μ_1, μ_2 be two probability measures on $\Gamma_{T,R,v_{\min}}$ and let $\gamma_k \sim \mu_k$, for $k = 1,2$. We take the same noise process W according to Model (12) in both cases: $W \sim \nu$ and W is independent from γ_1 and from γ_2. For the normalized functional $\bar{\rho}$ defined in (3) from on a kernel k that satisfies (4, 6, 7), if $p \geq 1 + \max(r_2, r_2/(r_1-1))$, then*

$$\|F(\phi\circ\gamma_1 + W) - F(\phi\circ\gamma_2 + W)\|_{\mathcal{H}} \leq \frac{\tilde{C}(K_{r_2,r_1})}{v_{\min}^{\alpha}} W_{1,\|\cdot\|_\infty}(\mu_1,\mu_2)^{\alpha},$$

where $\tilde{C}(x) = O(x^{1/r_2}(1 + x^{1/(r_1-1)}))$ depends on ϕ, ϵ, p, q and k.

Remark 3.5. Proposition 1.11 in Azaïs and Wschebor (2009) shows that if W satisfies (16), then W has a version with α-Hölder continuous sample paths, for any $0 < \alpha < r_1/r_2$. Difficulties in treating W come both from controlling its amplitude and the regularity. The tools that we use are sensitive to many, small fluctuations. Condition (16) allows us to control the regularity, without imposing a uniform Hölder character on all paths.

Sketch of the proof. The idea of the proof is to treat S path-wise and bound $\|\bar{\rho}(\phi\circ\gamma_1 + W) - \bar{\rho}(\phi\circ\gamma_2 + W)\|_{\mathcal{H}} \leq C\|W_{\gamma_1^{-1}} - W_{\gamma_2^{-1}}\|_\infty$. The constant C here is stochastic, but we can bound its moments using the regularity of total ϵ-truncated p-persistence and that of W. Finally, a lower-bound on the modulus of continuity of γ allows us to upper-bound $\|W_{\gamma_1^{-1}} - W_{\gamma_2^{-1}}\|_\infty$ in terms of $\|\gamma_1 - \gamma_2\|_\infty / v_{\min}$. We defer the full proof of Theorem 3.4 to Section D.1 of the Supplementary Material.

We now discuss two cases showing that the control in Theorem 3.4 is satisfactory. First, suppose that $\mu_k = \delta_{\gamma_k}$ for $k = 1,2$, for some fixed $\gamma_1, \gamma_2 \in \Gamma_{T,R,v_{\min}}$. Then, we obtain that the silhouette is Hölder, with respect to the distance $\|\gamma_1 - \gamma_2\|_\infty$. It is expected that we do not have complete invariance: for a fixed path W, the reparametrisation γ can influence how the points in the persistence diagram are displaced. Consider now the case of vanishing noise. If K_{r_2,r_1} decreases to zero, then so does the Hölder constant Λ_W and we have indeed that the right-hand side becomes zero.

Note that controlling $\|W\|_\infty$ is not sufficient for the stability. When $A_W < \epsilon$, the constant factor in $\tilde{C}(x)$ is $C_{\Lambda_W} = L_k(1 + 8p^2 A_\phi(A_\phi - \epsilon)\text{pers}_{p-2,\epsilon}^{p-2}(\phi)/(R-2)q^p)$. We can take the truncation parameter ϵ small, in which case $q \approx (A_\phi - \epsilon)$ and so, for a function with a single maximum and minimum, we have $C_{\Lambda_W} \approx L_k(1 + 8p^2) > 0$, which is not zero. Even though the amplitude of the noise is smaller than the cut-off ϵ, it still has an influence on the signature. Therefore, it is important that as the amplitude decreases, the noise does not become increasingly irregular: it is the case of aW, with $a \to 0^+$. The almost-sure bound on A_W gives us the lower–bound on $\text{pers}_{p,\epsilon}^{p}(\phi\circ\gamma + W)$, which appears in the denominator of $\bar{\rho}$.

For processes of decreasing amplitude but increasingly irregular, it is more advantageous to bound $\|W_{\gamma_1^{-1}} - W_{\gamma_2^{-1}}\|_\infty \leq 2\|W\|_\infty$ in the proof. In such a scenario however, we ignore the reparametrisations

so the distance $\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty$ disappears from the bound. Finally, it would be interesting to extend Theorem 3.4 to $\Gamma_{v_{\min}}$ from Proposition 3.3. In that case, we do obtain $W_1((\mu_1)_x, (\mu_2)_x)$ for almost-all $x \in \mathbb{R}^2$, but it is not clear that it lower–bounds $W_1(\mu_1, \mu_2)$.

Remark 3.6. When both endpoints are fixed and common to all reparametrisations, there is no reason to normalize by the total persistence. The stability comes from the continuity of the functional, not the renormalisation. Proposition 2.10 states that linear functionals of the form $\sum_{x\in D} w_\epsilon(x)^p k_x$ are also continuous for Hölder functions, so a statement analogue to Theorem 3.4 also holds for such functionals.

Remark 3.7. A similar result to Proposition 2.14 could be shown for the signature F, using the regularity assumption (16).

3.3. Estimation of the signature of a time series

We have defined the signature and studied its properties for continuous observations. In practical applications, we do not have access to S, but to observations in the form of a time series $\mathbf{X} = (X_n)_{n=1}^N$. The purpose of this section is to show how to exploit the periodicity mechanism to obtain asymptotic statistical guarantees for signatures of a discretized signal.

3.3.1. Time series model

We now consider a time series $(X_n)_{n\geq 1}$ which appears as a reparametrisation of a 1-periodic function ϕ:

$$X_n = \phi(\gamma_n) + W_n \in \mathbb{R}, \quad n \geq 1, \tag{17}$$

where $(\gamma_n)_{n\geq 1}$ is a strictly increasing time series and $(W_n)_{n\geq 1}$ is a stationary noise time series satisfying $\mathbb{E}[W_n] = 0$ and $|W_n| \leq (A_\phi - \epsilon - q)/2$ almost-surely. Moreover, $(\gamma_n)_{n\geq 1}$ and $(W_n)_{n\geq 1}$ are assumed to be independent.

We consider a class of reparametrisation processes, defined as discrete integrals of another, positive time series V_n. Specifically, let

$$\gamma_{n+1} = \gamma_n + hV_n = \gamma_0 + h\sum_{k=0}^{n} V_k, \tag{18}$$

where $(V_n)_{n=0}^N$ is a sequence of random variables in $I := [v_{\min}, v_{\max}] \subset]0, \infty[$, independent of γ_0 and $h > 0$ is a time step.

This model is inspired by dynamics, where the sequence $(\gamma_n)_{n\in\mathbb{N}}$ could model the displacement of a body over time and V_n should be thought of as the instantaneous speed. As for the continuous framework studied before, the latent times γ_n are not assumed to be observed.

We will consider two models for $(V_n)_{n\in\mathbb{N}}$. In the first one, consecutive velocities are independent. Since we do not expect a moving body to change speed abruptly, we also consider V_n as a Markov process on I. Basically, we need the velocity in these models to "see" enough data for the statistical theory presented later to work, thanks to the periodicity of the underlying mechanism.

Model 1 ($(V_n)_{n\in\mathbb{N}}$ i.i.d). We assume that V_n are independent and follow the same, unknown distribution on $\mathbb{R}_+^*$, which satisfies the following property: there exists $0 < a, b, c$ such that, for all $A \in \mathcal{B}(]a, b[)$ measurable, $P(V_k \in A) \geq c\mu(A)$, where μ is the Lebesgue measure.

Model 2 ($(V_n)_{n\in\mathbb{N}}$ a Markov Chain). Let $(V_n)_n$ be a Markov Chain with transition kernel P. Specifically, $v \mapsto P(v, A)$ is $\mathcal{B}(I)$-measurable for all $A \in \mathcal{B}(I)$, and $A \mapsto P(v, A)$ is a probability measure on $(I, \mathcal{B}(I))$. We further assume that $P(x, \cdot)$ is a probability measure that has a density f_x with respect to μ and that:

1. the density is lower–bounded in a small neighborhood: there exists $\eta, \mu_0 > 0$, such that
$$f_v|_{[v-\eta, v+\eta]\cap I} \geq \mu_0, \tag{19}$$
2. $v \mapsto f_v(x)$ is continuous for any $x \in I$.

Note that if $f_x = f$, for all $x \in I$, Model 2 reduces to a particular case of the i.i.d setting, where P has density f, $a = v_{\min}, b = v_{\max}$ and $c = \mu_0$.

Example 3.8. Set $V_0 \sim \mathcal{U}(I)$ and let $0 < \eta < (v_{\max} - v_{\min})/4$. An example of a transition kernel satisfying assumptions of Model 2 is a truncated Gaussian kernel. The truncation is such that the support is I and $\sigma = \eta$.

3.3.2. Stationary regime for Models 1 and 2

In the next section, we will introduce a topological signature for $(X_n)_{n\geq 1}$. To do this, we consider the stationary regime, as we assumed earlier for continuous processes. We will also discuss extensions to the non stationary setting at the end of the section. Note that $(\gamma_n)_{n\geq 1}$ in (18) is not generally a stationary time series. For instance, for both Models 1 and 2, we have $P(\gamma_n < \gamma_{n+1}) = 1$.

For Model 1, the crucial observation is that the time series $(\operatorname{frac}\gamma_n)_{n\geq 1}$ is stationary, where $\operatorname{frac} x = x - \lfloor x \rfloor$ denotes the fractional part of a real number. See Appendix E.1 for a formal proof. The time series $(\phi(\gamma_n))_{n\geq 1} = (\phi(\operatorname{frac}\gamma_n))_{n\geq 1}$ and $(X_n)_{n\geq 1}$ are thus stationary.

Regarding Model 2, it can be checked that $(\operatorname{frac}\gamma_n, V_n)_{n\geq 1}$ is a Markov Chain which admits a stationary measure π_2 (see Appendix E.2 for details). Under the initial measure π_2 the time series $(\operatorname{frac}\gamma_n)_{n\geq 1}$ is thus stationary and so is $(X_n)_{n\geq 1}$. We will thus assume this initial measure in order to stay in the stationary setting and denote *Model 2-S* this setting.

3.3.3. Signature of the time series

As we did for the continuous model, we operate at a fixed time horizon. In the time series setting, we fix $M \in \mathbb{N}$. From the observed time series $(X_n)_{n=1}^N$, we define "windows" $\mathbf{X}_n$, each of length M along the time series by

$$\mathbf{X}_n = (X_n, \dots, X_{n+M-1}) \qquad \text{for } n = 1, \dots, N - M + 1. \tag{20}$$

Each vector $\mathbf{X}_n$ is a $\mathbb{R}^M$-valued random vector, and under Models 1 and 2-S, the time series $(\mathbf{X}_n)_{n\geq 0}$ is stationary.

We want to extend the definition of the signature F to compute it on the $\mathbf{X}_n$'s. Starting from a kernel k, we first extend the definition of the normalized functional $\bar{\rho}$ to define it on a vector of length M by

$$\bar{\rho}(\mathbf{X}_1) := \bar{\rho}(\tilde{S}_M), \tag{21}$$

where $\tilde{S}_M$ is a continuous process on $[0, T]$ which interpolates between entries of $\mathbf{X}_1$. Specifically, we define $\tilde{S}_M$ by prescribing its values on the set $((m-1)T/(M-1), X_m)_{m=1}^M$ and linearly interpolating in between. The resulting process $\tilde{S}_M$ follows the continuous model (12), so that $\bar{\rho}(\tilde{S}_M)$ is well-defined. We then define the signature by integrating $\bar{\rho}(\mathbf{X}_1)$:

$$F_M(\mathbf{X}_1) := \mathbb{E}[\bar{\rho}(\mathbf{X}_1)].$$

Note that $F_M(\mathbf{X}_1)$ is a function from $\mathbb{U}$ to $\mathbb{R}$. Under Models 1 and 2-S, we obviously have $F(\mathbf{X}_1) = F(\mathbf{X}_n)$ for any $n = 1 \dots N - M + 1$, we thus can write

$$F_M := F_M(\mathbf{X}_1) = F_M(\mathbf{X}_n).$$

We now discuss the interest of introducing this signature for the study of time series. When the discrete model (17)-(18) corresponds to discrete observations of an underlying continuous model of the form of (1), then the interpolated signal $\tilde{S}_M$ will be close to its corresponding complete signal S, when ϕ and the noise process are smooth enough. In the noiseless case, Theorem 2.13 shows that $F(S)$ converges to $F(\phi_{[c,c+1]})$, which is an intrinsic signature of the underlying signal ϕ, as the number of observed periods tends to infinity. To summarize, in an idealized situation where there is very little noise, for $\mathbf{X}_1$ a dense enough sampling along the continuous signal, and for M large enough, the quantity $F_M(\mathbf{X}_1)$ will be close to the intrinsic quantity $F(\phi_{[c,c+1]})$ which does not depend on the parametrization. Out of the previous idealized setting, the signature $F_M(\mathbf{X}_1)$ also depends on the noise distribution. However, the signature is robust (as justified by Theorem 3.4 in the continuous setting). This signature then can be used for standard data sciences purposes, as for instance change point detection along a periodic phenomenon subject to phase variations, and for which we wish to develop indicators robust to phase variations.

3.3.4. Estimation of the signature

In practice, the signature F_M has to be estimated because the expectation is unknown. The empirical counter-part of F_M is the empirical mean

$$\hat{F}_{M,N} = \tfrac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(\mathbf{X}_n),$$

where $\bar{\rho}(\mathbf{X}_n)$ is defined according to (21). The distribution of $\hat{F}_{M,N}$ can estimated by Moving Block Bootstrap (MBB), see Bühlmann (2002). To be specific, let $L = L(N - M + 1) \in \mathbb{N}$ be the block length and let $B = \frac{L}{N-M+1}$ the number of blocks we now define, which, without loss of generality, is assumed to be integer-value. The MBB consists of sampling B blocks, each composed of L consecutive vectors $\mathbf{X}_i$: that is, $(\mathbf{X}_n, \dots, \mathbf{X}_{n+L-1})$, for $n \in \{1, \dots N - M + 1\}$. The MBB sample is then

$$(\mathbf{X}^*_1, \dots, \mathbf{X}^*_{N-M+1}) := (\underbrace{\mathbf{X}_{n_1}, \dots, \mathbf{X}_{n_1+L-1}}_{\text{block 1}}, \underbrace{\mathbf{X}_{n_2}, \dots, \mathbf{X}_{n_2+L-1}}_{\text{block 2}}, \dots, \underbrace{\mathbf{X}_{n_B}, \dots \mathbf{X}_{n_B+L-1}}_{\text{block B}}),$$

where $n_1, \dots n_B \sim \mathcal{U}(1, \dots, N - M + 1)$ are independent. We finally define the bootstrap signature

$$\hat{F}^*_{M,N} = \tfrac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(\mathbf{X}^*_n).$$

Note that the MBB strategy is applied to the time series of vectors $(\mathbf{X}_n)_{n=1}^{N-M+1}$, and not to the initial time series $(X_n)_{n=1}^{N}$. Also note that the bootstrap sample contains overlapping samples at two different levels. Not only are the windows $\mathbf{X}_1, \dots, \mathbf{X}_N$ overlapping, but also the different blocks can overlap.

We now prove that the empirical mean $\hat{F}_{M,N}$ converges to F_M and that we can approximate the distribution of $\hat{F}_{M,N}$ by that of the bootstrap signature $\hat{F}^*_{M,N}$, as $N \to \infty$. The core idea resides in exploiting the periodicity to control how the dependence between $\mathbf{X}_1$ and $\mathbf{X}_{1+k}$ changes as k increases. For this, we recall the definition of two mixing coefficients, see for instance Doukhan (1995).

For a stationary sequence $(Y_n)_{n\in\mathbb{N}}$ of random variables, denote by $\sigma_{a,b}$ the σ-algebra generated by $Y_a, \dots Y_b$. The k-th ϕ-mixing coefficient is

$$\phi_Y(k) = \sup_{A\in\sigma_{-\infty,0}, B\in\sigma_{k,\infty}, P(A)>0} |P(B|A) - P(B)|,$$

The k-th β-mixing coefficient is

$$\beta_Y(k) = \frac{1}{2} \sup_{\substack{\mathcal{A}\subset\sigma_{-\infty,0},\\ \mathcal{B}\subset\sigma_{k,\infty}}} \sum_{A\in\mathcal{A}, B\in\mathcal{B}} |P(A\cap B) - P(A)P(B)|,$$

where $\mathcal{A}$, $\mathcal{B}$ are countable partitions of the sample space. We say that $(Y_n)_{n\in\mathbb{N}}$ is *absolutely regular* (or β-mixing) if $\beta_Y(k) \to 0$ as $k \to \infty$. A process for which $\beta(k) \le a^k$, for some $0 < a < 1$ is called *exponentially β-mixing*. The same definitions apply for the uniform mixing coefficients ϕ_Y. Moreover we have that $\beta_Y(k) \le \phi_Y(k)$.

Theorem 3.9. *Let the stationary time series $(X_n)_{n\ge1}$ defined by (18) and (20), with $(V_n)_{n\ge1}$ as in Model 1 or 2-S. Assume that $(W_n)_{n\ge1}$ is exponentially β-mixing. Then,*

$$\sqrt{N-M+1}(\hat{F}_{M,N} - F_M) \to G_M \quad \text{as } N\to\infty, \tag{22}$$

where G_M is a zero–mean Gaussian process with covariance

$$(s,t) \mapsto \lim_{k\to\infty} \sum_{n=1}^{\infty} \operatorname{cov}\Big(\bar{\rho}(\mathbf{X}_k)(s), \bar{\rho}(\mathbf{X}_n)(t)\Big). \tag{23}$$

In addition, if $L(N) \to \infty$ and $L(N) = O(N^{1/2-\epsilon})$ for some $\epsilon > 0$, then the bootstrap is almost surely valid:

$$\sqrt{N-M+1}(\hat{F}^*_{M,N} - \hat{F}_{M,N}) \to^* G_M \quad \text{as } N\to\infty, \tag{24}$$

where the convergence is for the (conditional) bootstrap distribution.

This result is a functional central limit theorem, similar to many in the literature of topological data analysis, see for example Chazal et al. (2014) and Berry et al. (2018, Proposition 2 and 3), except that the samples are not independent. For independent data, it is sufficient to control the complexity of the functional family. The novel aspect of Theorem 3.9 is the consideration of dependence and it is what we treat with more care. More precisely, we show the following result:

Proposition 3.10. *Let the time series $(X_n)_{n\ge1}$ defined by (20) with $(V_n)_{n\ge1}$ as in (18).*

1. *If $(V_n)_{n\ge1}$ satisfies Model 1 then the Markov Chain $(\operatorname{frac}\gamma_n)_{n\ge1}$ is exponentially ϕ-mixing.*
2. *If $(V_n)_{n\ge1}$ satisfies Model 2 then the Markov Chain $(\operatorname{frac}\gamma_n, V_n)_{n\ge1}$ is exponentially ϕ-mixing, whatever the initial measure.*

Note that the proposition is valid for Model 2 in general, meaning also for non stationary regimes, and in that case more general definitions for β- and ϕ- mixing coefficients are used, see for instance Doukhan (1995).

The complete proof of Proposition 3.10 is given in Section F of the the Supplementary Material. It relies on a general theorem for Markov chains given in Doukhan (1995), and which requires a Doeblin-type condition.

Sketch of the proof of Proposition 3.10. For Model 1, we show that an uniform measure with small but non–zero mass lower-bounds the distribution of $\mathrm{frac}(\sum_{k=0}^{n} V_k)$. The fact that the process $(\mathrm{frac}(\gamma_0 + \sum_{k=0}^{n} V_k)_{n\in\mathbb{N}}$ is ϕ-mixing then follows from general results in dependence theory (Doukhan, 1995, Section 2.4, Theorem 1). When V_n is generated using Model 2, we find a similar uniform lower–bound for the n-step transition measure of $(\mathrm{frac}\,\gamma_n, V_n)$. With the assumptions on the kernel in our model, we show that for n sufficiently large, this lower–bound can be taken to be a uniform measure on $[0,1]\times I$ with small but non–zero mass, chosen uniformly in the initial conditions $(\mathrm{frac}\,\gamma_0, V_0)$. We conclude again with (Doukhan, 1995, Section 2.4, Theorem 1). This concludes the proof that $(\mathrm{frac}\,\gamma_n, V_n)_{n\in\mathbb{N}}$ is exponentially ϕ-mixing. □

We now give the proof of Theorem 3.9, with all details differed to the Supplementary Material.

Proof of Theorem 3.9. The functional Central Limit Theorem we use to show Theorem 3.9 relies on the control of β-mixing coefficients in the stationary regime. Moreover, we know that β-mixing coefficients are upper bounded by ϕ-mixing coefficients. Thus, under the assumptions of the theorem, we obtain from Proposition 3.10 that

- under Model 1: $(\mathrm{frac}\,\gamma_n)_{n\geq 1}$ is exponentially β-mixing.
- under Model 2-S: $(\mathrm{frac}\,\gamma_n, V_n)_{n\geq 1}$ is exponentially β-mixing, and thus the marginal $(\mathrm{frac}(\gamma_n))_{n\geq 1}$ is also exponentially β-mixing.

Next, we analyze how the dependence of $(\phi(\gamma_n))_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$ shapes the dependence of $(X_n)_{n\in\mathbb{N}}$ and that between the windows $\mathbf{X}_1,\ldots,\mathbf{X}_{N-M+1}$. Specifically, we have the following inequality

$$\beta_{\mathbf{X}}(k) \leq \beta_X(k-(M+1)) \leq \beta_{\mathrm{frac}(\gamma)}(k-(M+1)) + \beta_W(k-(M+1)), \qquad \text{for } k \geq M+1,$$

of which we present a detailed proof in Section G of the Supplementary Material. Since $(W_n)_{n\in\mathbb{N}}$ is exponentially β-mixing by assumption, $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is exponentially β-mixing. The Gaussian approximation (22) is a consequence of Kosorok (2008, Theorem 11.24), the statement of which is included in Section H of the Supplementary Material. Indeed, the arguments above ensure that the mixing coefficients satisfy

$$\sum_{k=1}^{\infty} k^{\frac{2}{r-2}} \beta_{\mathbf{X}}(k) < \infty,$$

for some $r \in]2,\infty[$. It remains to verify that the bracketing entropy of the functional family $\{\bar{\rho}_u\}_{u\in\mathbb{U}}$ is controlled, see Section I of the Supplementary Material.

The approximation of the distribution of the empirical mean by the bootstrap distribution (24) is a consequence of Bühlmann (1995, Theorem 1) , for which we only need the aforementioned results (see Section H of the Supplementary Material). □

Remark 3.11. Note that length M of the window is fixed in the Theorem, it does not vary with N. It would be interesting to make this quantity increases with N. For instance, when we use this signature to compare two time series, intuitively increasing M may improve the discriminative power of the signature. But too large M will also decrease the sample size of windows $\mathbb{X}_n$, and then the variance of the empirical signature will increase. Choosing M is a non-trivial issue, which moreover cannot be easily resolved in practice by a cross-validation approach.

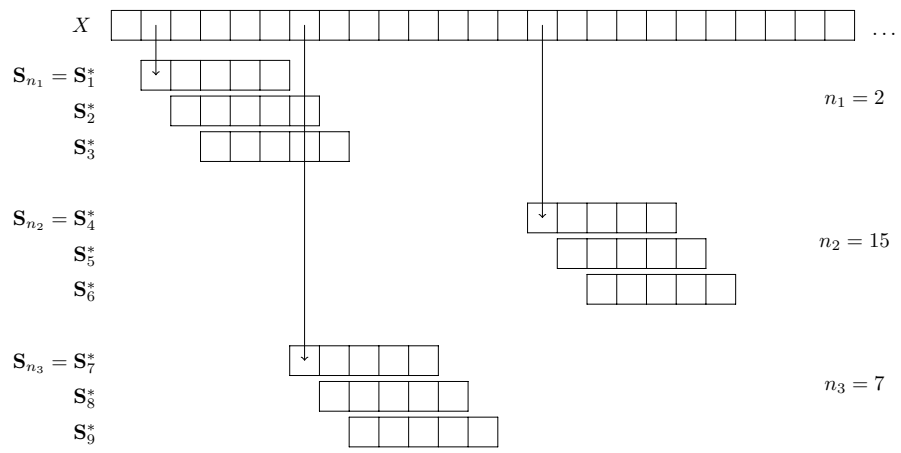

Figure 5. A schematic representation of the MBB, for $M = 5$ and $L = 3$.

Remark 3.12. A classical framework in Functional Data Analysis (FDA) is when we have access to a collection of K time series from the same model. Many contributions have been proposed to study this statistical setting, see for instance Ramsay and Silverman (2002). To recast our setting in this standard framework of FDA, we need to know how to segment the curve, or the time series, into successive periods, which is not trivial in practice when there is unknown reparametrization and phase variations in the signal, see for instance Bonis et al. (2022). The advantage of our approach is precisely that it avoids this additional step of segmentation.

3.3.5. Discussion on possible extensions

The literature on functional central limit theorems for dependent data is rich in results for various functional classes and dependence assumptions. We believe it might be possible to use more recent and stronger results than Bühlmann (1995, Theorem 1). This would allow us to relax the decay of β_W from an exponential to a polynomial one. For instance, Radulović (1996, Theorem 1) is written for VC-classes functionals, but the proof seems to rely on the bracketing entropy bound that the functionals considered in the present work also satisfy.

An other natural extension to Theorem 3.9 is the framework of non stationary time series for Model 2. Although the times series $(X_n)_{n\geq 1}$ is non stationary, we still define the topological signature F_M with as initial measure the invariant measure π_2. According to (Doukhan, 1995, Theorem 1, Section 2.4), the Markov Chain $(\text{frac}\,\gamma_n, V_n)_{n\geq 1}$ is ϕ-mixing and also uniformly ergodic. We can then show convergence results quite directly only based on Markov chain properties. To keep things simple, let's consider Model 2 with no noise ($W_n = 0$). Then, $(\text{frac}\,\gamma_n, V_n)$ admits a Central Limit Theorem (CLT), as well as $(\phi\,(\text{frac}\,\gamma_n)))_{n\geq 1}$, see for instance Theorem 24 and Proposition 29 in Roberts and Rosenthal (2004). From this we can deduce that $\hat{F}_{M,N}$ admits a pointwise CLT, meaning for fixed u.

If we now consider mixing properties of Model 2 (with additive noise W_n), several avenues of research are still possible to demonstrate CLTs under non stationary regime. For instance, a CLT for ϕ-mixing sequences was obtained in Utev (1990), by assuming the Lindeberg condition. In Rio (1997), a CLT for strong mixing is given that applies to both non-stationary sequences and triangular array settings (see also Rio et al., 2017). More recently, a functional CLT is proposed Merlevede and Peligrad

(2020) for non-stationary strongly mixing sequences. This last contribution certainly paves the way for demonstrating a result analogous to (22) in a non-stationary setting.

4. Numerical illustration

To illustrate the signatures and their stability, we propose to estimate the signatures of processes with different periodic functions. Then, we compare the estimate to the signature of a process with a different reparametrisations.

We will consider periodic functions ϕ_1 and ϕ_4 defined by

$$\phi_\theta = \theta(\sin(6\pi t) + |t - \lfloor t \rfloor - \tfrac{1}{2}| - \tfrac{1}{2}) + 5\sin(4\pi t), \qquad \text{for } \theta \in \mathbb{R}.$$

The observed signal follows the discrete model (17), with $T = 30$ and a sampling rate of 50Hz. The reparametrisations are generated by integrating twice a Markov chain of accelerations, with a truncated Gaussian transition kernel. The noise is a Gaussian process with covariance

$$\Gamma(s,t) = \sigma^2 \exp\left(-\frac{(s-t)^2}{2\tau^2}\right).$$

We fix the temporal scale τ, but we vary $\sigma = 0.1, 0.5, 2.$ to illustrate the impact of noise on the signature.

For $\bar{\rho}$, we take the silhouette introduced in Example 2.8, where the weights are the 0.2-truncated 1-persistence ($\epsilon = 0.2$, $p = 1$) and we use the projection $\pi_{-9,9}$ as in (8). We infer the signatures on 3-second windows ($M = 3 \cdot 50$). We construct the 1%-confidence intervals by resampling 200 times, with block lengths of 2 seconds ($L = 2 \cdot 50$).

In Figure 6, for the same random realization γ_1, we calculate the empirical signature $\hat{F}$ for ϕ_1 and ϕ_4, and estimate the corresponding confidence intervals for F. For low noise levels, the variance due to the number of observations and the variability in the endpoints is small, compared to the difference between the functionals. As the noise level increases, the observed function looses its recurrent appearance and the signatures become dominated by the noise.

Consider now two observations with the same periodic function ϕ_1, but different reparametrisations γ_1, γ_2. In Figure 7, we can see that for small values of noise, the signatures are close, what confirms their invariance to reparametrisation. It is worth noting that the signals contain different numbers of periods. For more noisy observations, the signatures lose the robustness.

Acknowledgments

The authors are grateful to numerous colleagues for the fruitful discussions and they wish to particularly thank Paul Doukhan, Giovanni Peccati, Alex Delalande, Quentin Mérigot and Daniel Perez.

Funding

WR was supported by TopAI ANR–19–CHIA–0001 and BM by GeoDSIC ANR-22-CE40-0007.

Supplementary Material

Proofs and details
This supplement contains proofs of this manuscript.

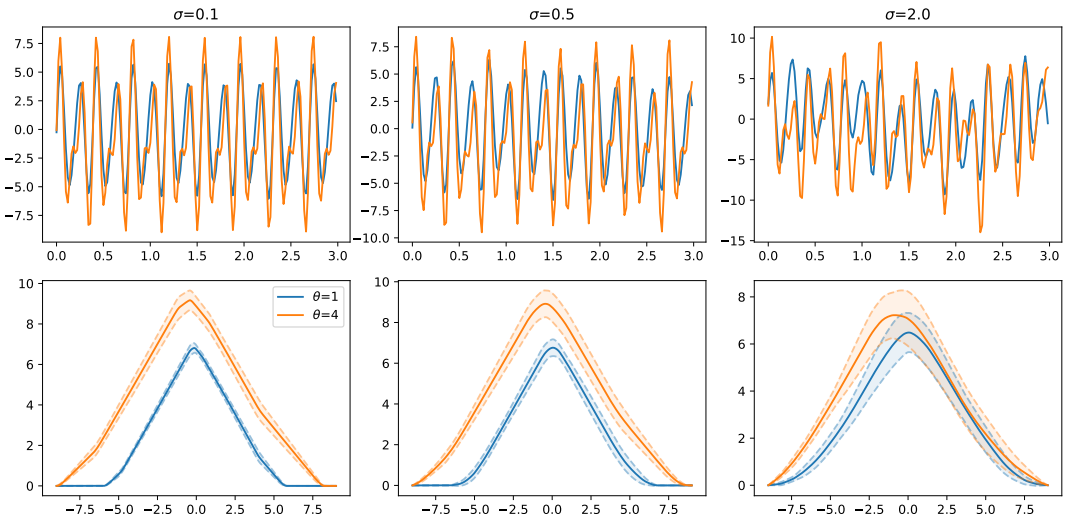

Figure 6. Signatures of ϕ_1 and ϕ_4, estimated on reparametrized signals described above. The top row shows the first 3-second window from the 30-second signal, for both functions. The bottom row shows the estimated signatures and the confidence intervals.

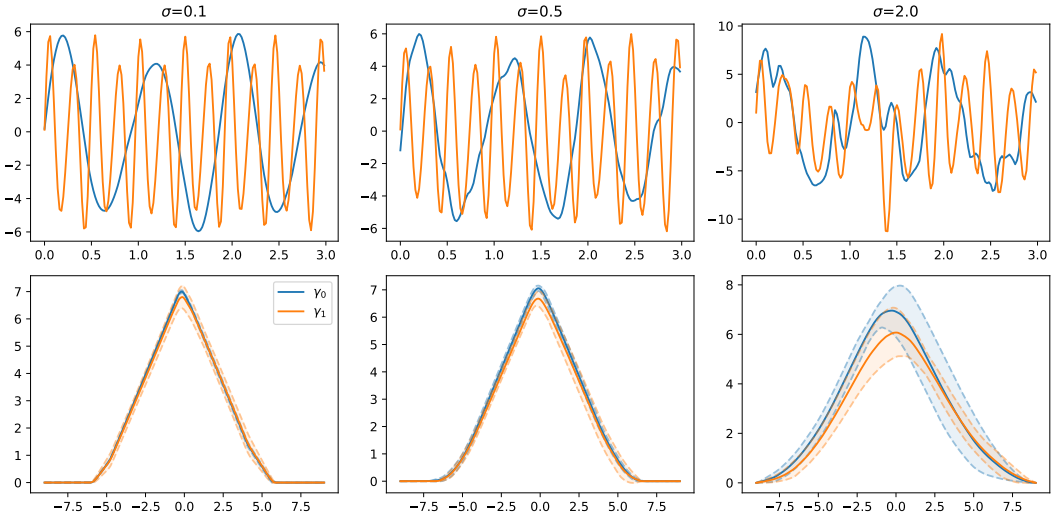

Figure 7. Signatures of ϕ_1, estimated on two different reparametrized observations. The top row shows the first 3-second window from the two observed signals. The bottom row shows the estimated signatures and the confidence intervals.

References

ADAMS, H., EMERSON, T., KIRBY, M., NEVILLE, R., PETERSON, C., SHIPMAN, P., CHEPUSHTANOVA, S., HANSON, E., MOTTA, F. and ZIEGELMEIER, L. (2017). Persistence Images: A Stable Vector Representation of Persistent Homology. *J. Mach. Learn. Res.* **18** 218–252.

AZAÏS, J.-M. and WSCHEBOR, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, Inc., Hoboken, NJ, Hoboken, NJ, USA.

BERRY, E., CHEN, Y.-C., CISEWSKI-KEHE, J. and FASY, B. T. (2018). Functional Summaries of Persistence Diagrams. *arXiv:1804.01618*.

BOGACHEV, V. I. (2007). *Measure Theory*. Springer, Berlin ; New York.

BOIS, A., TERVIL, B., MOREAU, A., VIENNE-JUMEAU, A., RICARD, D. and OUDRE, L. (2022). A Topological Data Analysis-Based Method for Gait Signals with an Application to the Study of Multiple Sclerosis. *PLOS ONE* **17**.

BONIS, T., CHAZAL, F., MICHEL, B. and REISE, W. (2022). Topological Phase Estimation Method for Reparameterized Periodic Functions. *Advances in Computational Mathematics* Accepted for publication.

BUBENIK, P. (2015). Statistical Topological Data Analysis using Persistence Landscapes. *J. Mach. Learn. Res.* **6** 77–102.

BÜHLMANN, P. (1995). The Blockwise Bootstrap for General Empirical Processes of Stationary Sequences. *Stochastic Process. Appl.* **58** 247–265.

BÜHLMANN, P. (2002). Bootstraps for Time Series. *Statist. Sci.* **17**.

CARRIÈRE, M., CHAZAL, F., IKE, Y., LACOMBE, T., ROYER, M. and UMEDA, Y. (2020). PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures. In *AISTATS'2017* **108** 2786–2796. PMLR.

CHAZAL, F. and MICHEL, B. (2021). An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front. Artif. Intell.* **4**.

CHAZAL, F., FASY, B. T., LECCI, F., RINALDO, A. and WASSERMAN, L. (2014). Stochastic Convergence of Persistence Landscapes and Silhouettes. In *SOCG'14* 474–483. ACM Press, Kyoto, Japan.

CHAZAL, F., DE SILVA, V., GLISSE, M. and OUDOT, S. (2016). *The Structure and Stability of Persistence Modules*. *SpringerBriefs in Mathematics* **2191-8198**. Springer, Cham.

CHUNG, Y.-M. and LAWSON, A. (2022). Persistence Curves: A canonical framework for summarizing persistence diagrams. *Adv. Comput. Math.* **48** 6.

COHEN-STEINER, D., EDELSBRUNNER, H. and HARER, J. (2007). Stability of Persistence Diagrams. *Discrete Comput. Geom.* **37** 103–120.

COHEN-STEINER, D., EDELSBRUNNER, H., HARER, J. and MILEYKO, Y. (2010). Lipschitz Functions Have L p -Stable Persistence. *Found. Comput. Math.* **10** 127–139.

CORCORAN, P. and JONES, C. (2017). Modelling Topological Features of Swarm Behaviour in Space and Time With Persistence Landscapes. *IEEE Access* **5** 18534-18544.

DEDECKER, J. (2007). *Weak dependence: with examples and applications*. *Lecture notes in statistics* **190**. Springer, New York.

DIVOL, V. and POLONIK, W. (2019). On the Choice of Weight Functions for Linear Representations of Persistence Diagrams. *J. Appl. Comput. Topol.* **3** 249–283.

DOUKHAN, P. (1995). *Mixing: Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer New York, NY.

FERNÁNDEZ, X. and MATEOS, D. (2022). Topological Biomarkers for Real-Time Detection of Epileptic Seizures.

GASSER, T. and WANG, K. (1997). Alignment of Curves by Dynamic Time Warping. *Ann. Statist.* **25** 1251–1276.

GHIL, M. and SCIAMARELLA, D. (2023). Review Article: Dynamical Systems, Algebraic Topology, and the Climate Sciences Preprint, Bifurcation, dynamical systems, chaos, phase transition, nonlinear waves, pattern formation/Climate, atmosphere, ocean, hydrology, cryosphere, biosphere/Theory.

GIDEA, M. and KATZ, Y. (2018). Topological Data Analysis of Financial Time Series: Landscapes of Crashes. *Phys. A* **491** 820–834.

GOLDBERGER, A. L., AMARAL, L. A., GLASS, L., HAUSDORFF, J. M., IVANOV, P. C., MARK, R. G., MIETUS, J. E., MOODY, G. B., PENG, C. K. and STANLEY, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101** E215-220.

KHASAWNEH, F. A. and MUNCH, E. (2016). Chatter Detection in Turning Using Persistent Homology. *Mech. Syst. Signal Process.* **70–71** 527–541.

KHORRAM, S., MCINNIS, M. G. and PROVOST, E. M. (2019). Trainable Time Warping: Aligning Time-Series in the Continuous-Time Domain. In *ICASSP'2019* 3502–3506.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer New York, New York, NY.

KREBS, J. (2021). On Limit Theorems for Persistent Betti Numbers from Dependent Data. *Stochastic Process. Appl.* **139** 139–174.

MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional Data Analysis of Amplitude and Phase Variation. *Statist. Sci.* **30** 468–484.

MERLEVEDE, F. and PELIGRAD, M. (2020). Functional CLT for nonstationary strongly mixing processes. *Stat Probabil Lett* **156** 108581.

PANARETOS, V. M. and ZEMEL, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. *SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham, Cham.

PEREA, J. A. (2019). Topological Time Series Analysis. *Notices Amer. Math. Soc.* **66** 1.

PEREZ, D. (2022). On C0-persistent homology and trees. 41.

PERNG, C. S., WANG, H., ZHANG, S. R. and PARKER, D. S. (2000). Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases. In *Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073)* 33–42. IEEE Comput. Soc., San Diego, CA, USA.

PLONKA, G. and ZHENG, Y. (2016). Relation between total variation and persistence distance and its application in signal processing. *Adv. Comput. Math* **42** 651–674.

RADULOVIĆ, D. (1996). The Bootstrap for Empirical Processes Based on Stationary Observations. *Stochastic Process. Appl.* **65** 259–279.

RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, 1 ed. *Springer Series in Statistics*. Springer-Verlag, New York, New York.

REISE, W. (2023). Topological techniques for inference on periodic functions with phase variation, PhD thesis, Université Paris-Saclay.

RIO, E. (1997). About the Lindeberg method for strongly mixing sequences. *ESAIM-Probab Stat* **1** 35–61.

RIO, E. et al. (2017). *Asymptotic theory of weakly dependent random processes* **80**. Springer.

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys* **1** 20 – 71.

SU, J., KURTEK, S., KLASSEN, E. and SRIVASTAVA, A. (2014). Statistical Analysis of Trajectories on Riemannian Manifolds: Bird Migration, Hurricane Tracking and Video Surveillance. *Ann. Appl. Stat.* **8** 530–552.

TANG, R. and MULLER, H. G. (2008). Pairwise Curve Synchronization for Functional Data. *Biometrika* **95** 875–889.

UTEV, S. A. (1990). *Central Limit Theorem For Dependent Random Variables* In *Vol. 2* 519–528. De Gruyter, Berlin, Boston.