

Persistence Barcodes versus Kolmogorov Signatures: Detecting Modes of One-Dimensional Signals*

Ulrich Bauer¹, Axel Munk^{2,3}, Hannes Sieling², and Max Wardetzky⁴

¹Technische Universität München (TUM)

²Institute for Mathematical Stochastics, University of Göttingen

³Max Planck Institute for Biophysical Chemistry, Göttingen

⁴Institute of Numerical and Applied Mathematics, University of Göttingen

19 August 2015

Abstract

We investigate the problem of estimating the number of modes (i.e., local maxima)—a well known question in statistical inference—and we show how to do so without presmoothing the data. To this end, we modify the ideas of persistence barcodes by first relating persistence values in dimension one to distances (with respect to the supremum norm) to the sets of functions with a given number of modes, and subsequently working with norms different from the supremum norm. As a particular case we investigate the *Kolmogorov norm*. We argue that this modification has certain statistical advantages. We offer confidence bands for the attendant *Kolmogorov signatures*, thereby allowing for the selection of relevant signatures with a statistically controllable error. As a result of independent interest, we show that *taut strings* minimize the number of critical points for a very general class of functions. We illustrate our results by several numerical examples.

AMS subject classification: Primary 62G05, 62G20, 68U05; secondary 62H12, 57R70, 58E05

Keywords: Persistent homology, mode hunting, exponential deviation bound, partial sum process, taut strings

Communicated by Shmuel Weinberger

1 Introduction

Persistent homology [16, 17] provides a *quantitative* notion of the stability or robustness of critical values of a (sufficiently nice) real valued function f on a topological space: the persistence of a critical value is a lower bound on the amount of perturbation (in the supremum norm) required for its elimination. Persistence measures the life span of homological features in terms of the difference between *birth* and *death* of such features—according to the filtration of the underlying

*Research partially supported by DFG FOR 916, Volkswagen Foundation, and the Toposys project FP7-ICT-318493-STREP

topological space that arises from the sublevel sets of f . Birth and death of homological features of f can be encoded in a barcode diagram, see [17]. In this article, we consider what we call *persistence signatures*, defined as (half) the life span (or persistence) of critical values, i.e., persistence signatures correspond to (half) the lengths of persistence barcodes and (when properly ordered) give rise to a descending sequence

$$s_{0,\infty}(f) \geq s_{1,\infty}(f) \geq s_{2,\infty}(f) \geq \cdots, \quad (1)$$

where we appropriately account for multiplicity of critical values. In our setup, $s_{0,\infty}(f)$ denotes the largest *finite* persistence value of f , and we append the sequence by zeros beyond the smallest positive persistence value of f .

We consider one dimensional signals $f : [0, 1] \rightarrow \mathbb{R}$. For the moment, to illustrate our results, let X denote the space of piecewise constant real-valued functions on a (variable) equipartition of $[0, 1]$. (Later in our exposition, we also consider more general function spaces.) Let $X_k \subset X$ denote the space of functions with at most k *modes*, i.e., local maxima, where we only count inner local maxima. Our point of departure is the observation that

$$s_{k,\infty}(f) = \text{dist}_\infty(f, X_k),$$

i.e., $s_{k,\infty}(f)$ equals the distance of f to the space of functions with at most k modes with respect to the sup norm. This follows from the combination of two facts. First, from the celebrated stability theorem in persistent homology [10], which asserts that

$$|s_{k,\infty}(f) - s_{k,\infty}(g)| \leq \|f - g\|_\infty \quad \text{for all } k \geq 0.$$

Second, from the fact that in order to eliminate all positive persistence signatures of $f : [0, 1] \rightarrow \mathbb{R}$ with value less or equal to δ , it suffices to change f by δ in the sup norm, see [2].¹

The fact that persistence signatures correspond to distances (with respect to the sup norm) to sets of functions with at most k modes leads us to considering norms different from the sup norm. Our motivation is to ask how signatures arising from different norms compare in a statistical sense. To this end, consider an arbitrary metric d on X and define the metric signatures

$$s_k(f) := \text{dist}(f, X_k) \quad \text{with respect to } d.$$

Then $(s_k(f))$ is an descending sequence as in (1), see Fig. 1. Moreover, since distance to sets in metric spaces is 1-Lipschitz, stability is immediate:

$$|s_k(f) - s_k(g)| \leq d(f, g) \quad \text{for all } k \geq 0.$$

The resulting signatures $s_k(f)$ will in general be different from persistence signatures. The aim of this article is to analyze, from a statistical and algorithmic point of view, one particular example: the *Kolmogorov metric* d_K and its resulting Kolmogorov signatures s_K . For one dimensional signals $f : [0, 1] \rightarrow \mathbb{R}$ the Kolmogorov norm is defined as the L^∞ -norm of the antiderivative F of f , subject to $F(0) = 0$. The Kolmogorov norm plays a prominent role in probability and statistics, see, e.g., [29]. Our approach is based on the observation that if $s_k(f) = 0$, then (the unknown function) f has at most k modes. This provides a link between *mode hunting*, a widely studied problem in statistics [22, 19, 23, 12, 30], and the robust estimation of signatures. Most related to

¹Note that this result does no longer hold in dimensions greater than two.

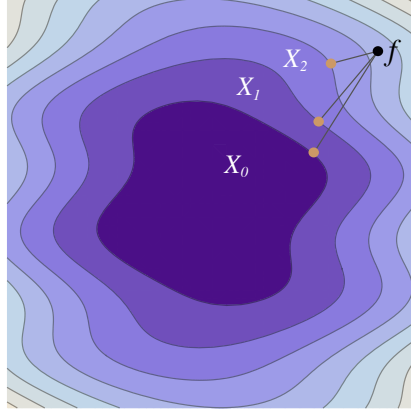


Figure 1: Illustration of metric signatures, i.e., distances of some $f \in X$ to the sets X_k containing those functions with at most k modes.

our approach is [12], where the Kolmogorov norm has been used for mode hunting in the context of density estimation.

In the sequel we consider the following basic statistical additive regression model. Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is perturbed by random noise ϵ and observed by a finite number of (equidistantly sampled) measurements $(Y_i)_{i=0}^n$, i.e.,

$$Y_i = f(t_i) + \epsilon_i, \quad t_i = i/n. \quad (2)$$

Throughout we assume that the noise (ϵ_i) is independently distributed with mean zero such that for some $\kappa > 0$, $v > 0$ and all $m \geq 2$,

$$\mathbb{E}[|\epsilon_i|^m] \leq vm! \kappa^{m-2}/2 \quad \text{for all } i = 1, \dots, n. \quad (3)$$

Distributions which satisfy this condition include the centered normal distribution, the (centered) Poisson distribution, or the Laplace distribution. Moreover, any symmetric distribution around zero with compact support is covered, including the uniform distribution on an interval $[-B, B]$.

We are concerned with the following question: *With what probability can one estimate the number of modes of f (or provide bounds for its under- and overestimation) from the observations Y ?*

In dimension one, where mode hunting is intimately related to persistent homology, this question has been addressed in topological data analysis (TDA). A well known problem in this context is the fact that the stability theorem of persistent homology is based on the sup norm, which potentially makes this approach non-robust to outliers or unbounded noise. Therefore, several methods have been recently suggested to overcome this problem in various settings [1, 3, 6, 5, 8, 9, 25, 28]. Roughly speaking, these methods have in common that they first regularize or filter the data in one form or another—in order to improve stability with respect to the sup norm—and then work with the persistence diagram of the so obtained preprocessed result. This is based on the initial estimation of f itself. From a statistical perspective, however, having to estimate f in a first step somewhat weakens the potential appeal of TDA. Already in dimension one of the underlying space, estimating f by any regularization technique leads to difficult problems, e.g., data driven smoothing or parameter thresholding. We stress that in

addition, this sensibly affects the resulting persistence properties in a statistically hard to control manner, see, e.g., [1, 5, 18] for the case of a kernel estimator. In fact, presmoothing with a kernel estimator leads to what has been sometimes called the *notorious bandwidth selection problem*, which does not possess a widely accepted solution since the optimal bandwidth (e.g., in the sense of minimizing the mean squared error between f and its kernel estimate)—although theoretically known—depends on unknown characteristics of f , such as its curvature (see [32] among many others). Hence, we argue that a conceptual simplification and a computational advantage of TDA would result from circumventing explicit estimation of f .

One aim of this paper is to show that direct estimation of topological properties of f *without* having to estimate f itself is indeed a doable task by using Kolmogorov signatures. We confine ourselves to dimension one because using the Kolmogorov norm in this case lends itself to an efficient algorithm ($O(n \log n)$ in the number of data points). We stress that our statistical analysis carries over higher dimensions.

A second aim of this paper is to provide confidence statements on the empirical Kolmogorov signatures with a controllable statistical error, similar in spirit to [18], where asymptotic confidence bands for the empirical (sup norm based) persistence diagram are given for data on a manifold. Their approach, however, is based on presmoothing for unbounded noise using a kernel density estimator, which we avoid in this paper.

Inference for Kolmogorov signatures Using the Kolmogorov metric and the resulting Kolmogorov signatures, we investigate how well the empirical signatures $s_k(Y)$, obtained by interpreting Y as a piecewise constant function, estimate the signatures $s_k(f)$. As a starting point, Theorem 1 asserts that under the moment condition (3), for any $\delta > 0$ one has

$$\mathbb{P}\left(\max_{k \in \mathbb{N}_0} |s_k(Y) - s_k(f)| \geq \delta\right) \leq 2 \exp\left(-\frac{\delta^2 n}{2\nu + 2\kappa\delta}\right).$$

Using this, Theorem 2 asserts that for a given probability $\alpha \in (0, 1)$, one can construct non-asymptotic confidence regions for the entire sequence $(s_k(f))$ of signatures in the sense that

$$\mathbb{P}(s_k(f) \in [(s_k(Y) - \tau_n(\alpha))_+, s_k(Y) + \tau_n(\alpha)] \text{ for all } k \in \mathbb{N}_0) \geq 1 - \alpha, \quad (4)$$

where $(x)_+ = \max(0, x)$. Here $\tau_n(\alpha)$ depends in an explicit manner on n , α , κ , and ν , which are known constants or can be easily estimated from the data. We drop the dependence of τ_n on κ and ν by considering κ and ν fixed since we are mainly concerned with the dependence on n and α . For fixed α, κ, ν , one asymptotically has $\tau_n(\alpha) \approx 1/\sqrt{n}$. The parameter $\tau_n(\alpha)$ can be used to threshold the empirical signatures $s_k(Y)$ by defining

$$k_\epsilon(Y) = \max\{k \in \mathbb{N}_0 : s_{k-1}(Y) \geq \epsilon\},$$

where, as a convention, we define $s_{-1}(Y) = \infty$. Then Theorem 3 asserts that for all $k \in \mathbb{N}_0$, $f \in X_k$, and $\alpha \in (0, 1)$, one has

$$\mathbb{P}(k_{\tau_n(\alpha)}(Y) > k) \leq \alpha,$$

i.e., the threshold parameter $\tau_n(\alpha)$ controls the probability of *overestimating* the number of modes for any function $f \in X$. Notice that $\tau_n(\alpha)$ is independent of the number and magnitude of the modes of f , so in this sense the result is universal. Obtaining a universal result in

the other direction, i.e., controlling the probability of *underestimating* the number of modes, is a more delicate task. Indeed, as pointed out in [14], obtaining such results is in general impossible if the modes of f are allowed to become arbitrarily small. As a consequence, without a priori information on the “smallest scales” of f , no method can provide a control for their underestimation. Therefore, it is *only* possible to provide a bound for underestimating those signatures of f that are larger than a certain threshold. Theorem 4 asserts that for any $k \in \mathbb{N}_0$, $f \in X_k$, and $\alpha \in (0, 1)$, one has

$$\mathbb{P}(k_{\tau_n(\alpha)}(Y) < k_{2\tau_n(\alpha)}(f)) \leq \alpha .$$

Combining the latter results, we obtain *two sided bounds* for the estimated number of modes. More precisely, for any $f \in X_k$ and any $\alpha \in (0, 1)$ we obtain that

$$\mathbb{P}(k_{2\tau_n(\alpha/2)}(f) \leq k_{\tau_n(\alpha/2)}(Y) \leq k) \geq 1 - \alpha .$$

As mentioned before, for fixed α, κ, v , one has $\tau_n(\alpha) \approx 1/\sqrt{n}$. Therefore there exists a constant C such that asymptotically (for large enough n) by thresholding at $C/\sqrt{n}$, it can be guaranteed at a level α that all signatures above this threshold are detected. Notice that so far we have not made use of any a priori information about f . This changes with Theorem 5, which asserts that if $f \in X_k$ and $s_{k-1}(f) \geq \epsilon$, then

$$\mathbb{P}(k_{\epsilon/2}(Y) = k) \geq 1 - 2 \exp\left(-\frac{\epsilon^2 n}{8v + 4\kappa\epsilon}\right) , \quad (5)$$

i.e., the number of modes of f can be estimated correctly with a probability that approaches 1 exponentially fast by thresholding the empirical signature *provided that* one has a priori lower bounds on magnitude (in the Kolmogorov norm) of the smallest mode of f . Notice that this result is independent of the number of modes of f .

Kolmogorov signatures vs. persistence signatures Kolmogorov signatures offer an alternative to persistence signatures, since they behave more robust for large errors ϵ_i . The intuitive reason is that the Kolmogorov norm damps these errors, while they remain dominant using the sup norm without prefiltering. This is relevant, e.g., for unbounded noise (such as normally distributed errors, which are included in our noise model (3)) or for data with outliers. Additionally, our results also indicate how inference for persistence signatures becomes increasingly difficult with decreasing signal to noise ratio even in the case of bounded noise. This is in accordance with, e.g., [18], where the noise bound explicitly enters into the bound of the estimation error for the correct number of modes.

Nevertheless, Kolmogorov signatures are not always superior to persistence signatures in terms of statistical efficiency. This can be seen by comparing their probabilities to detect a non vanishing signature from the data. To this end, we consider two limiting scenarios. The first comprises sparse signals with high peaks and small support, while the second comprises weak signals with large support. To illustrate these scenarios, we consider functions with one *single* mode and i.i.d. normal errors with variance one, i.e., $\epsilon_i \sim \mathcal{N}(0, 1)$.

In the first scenario, we consider a sequence of functions

$$f_n(x) = \begin{cases} (1 + \epsilon) \sqrt{2 \log n} & \text{if } x \in [j/n, (j+1)/n) , \\ 0 & \text{otherwise} , \end{cases} \quad (6)$$

for some $\varepsilon > 0$ and for some $j \in \{0, \dots, n-1\}$ that is *a priori not known*. We show in Theorem 6 that asymptotically (as $n \rightarrow \infty$) it is impossible to distinguish f_n from the zero function by thresholding Kolmogorov signatures at $\tau_n(\alpha)$ as above. In contrast, for such signals, sup norm based thresholding of the vector $(Y_1, \dots, Y_n)$ is known to behave asymptotically minimax efficient in the sense of detecting a non vanishing mode with probability tending to one as $n \rightarrow \infty$, see, e.g., [13, 24]. Whether this efficiency carries over to persistence signatures is unknown to us.

In the second scenario, we consider a sequence of functions

$$f_n(x) = \begin{cases} \delta_n & \text{if } x \in [1/3, 2/3), \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

with $\delta_n \rightarrow 0$. It is well known that it is possible to detect the single mode of f_n with probability tending to one as $n \rightarrow \infty$ if $\delta_n \sqrt{n} \rightarrow \infty$, see, e.g., [31]. From (5) it follows, using $\epsilon = \delta_n$, that Kolmogorov signatures can correctly detect the single mode of signals in (7) by thresholding signatures at $\delta_n/2$. In contrast, for persistence signatures, there exists no thresholding strategy that can detect the single mode with probability one. To be precise, let again $\delta_n \sqrt{n} \rightarrow \infty$, and assume additionally $\delta_n \sqrt{\log n} \rightarrow 0$. Then Theorem 7 asserts that for an arbitrary sequence (q_n) of reals one has $\limsup_{n \rightarrow \infty} \mathbb{P}(k_{q_n}(Y) = 1) < 1$.

While it might be intuitively clear that in general there exists no thresholding strategy for persistence signatures in the case of unbounded noise, we would like to add three remarks. First, we would like to remind the reader that persistence signatures in combination with regularization have been successfully applied to the problem of estimating topological features. However, this approach essentially comes at the above mentioned price of having to estimate the signal f itself, which is problematic from a statistical perspective, and which we avoid here. Second, apart from our Theorem 7, we are not aware of a proof that a thresholding strategy for persistence signatures without regularization cannot exist in general. Third, and perhaps surprisingly, for some signals a thresholding strategy for persistence signatures is indeed possible even in the case of unbounded noise. To illustrate the latter, consider a piecewise constant function f such that $f(x) = 1$ if $x \in [1/3, 2/3)$ and $f(x) = 0$ else. Then (under the assumption of i.i.d. normal noise) extreme value theory informs us that with a probability approaching one (as the number of observations tends to infinity) the smallest values in the data (Y_i) on both intervals $[0, 1/3)$ and $[2/3, 1]$ are approaching $-\sqrt{2 \log n}$; likewise, the largest values in (Y_i) on $[1/3, 2/3)$ are approaching $1 + \sqrt{2 \log n}$ with a probability going to one. As a consequence, if one thresholds the persistence signatures of the data at $\sqrt{2 \log n}$, then with a probability going to one, it is indeed possible to detect the single mode of f .

Efficient computation using taut strings While our approach can in principle be extended to metric different from the ones induced by the sup or Kolmogorov norms, not every metric lends itself to an efficient computation of the requisite signatures. The difficulty is to compute the distance of a given function to the set of functions with at most k modes. Using *taut strings* (which are intimately related to total variation (TV) minimization [20, 21, 11, 26]), we prove that the set of Kolmogorov signatures can be computed in $O(n \log n)$ time, where n is the number of observations. Given $f \in L^\infty([0, 1])$ and $\alpha \geq 0$, the *taut string*, F_α , is the function whose graph has minimal total length (as a curve) among all absolutely continuous functions in the α -tube around the antiderivative F of f . Letting $f_\alpha = F'_\alpha$ denote the derivative of the taut string, Theorem 8 provides a result of independent interest that has been implicitly used several times in the existing literature but has never been proven rigorously to our knowledge: f_α minimizes the number of

modes among all L^∞ -functions in the (closed) α -ball around f with respect to the Kolmogorov norm. Indeed, our result generalizes previous results on the mode-minimizing property of f_α , which were shown in the special context of piecewise constant functions using the Kolmogorov norm, see, e.g., [11, 12, 23, 26].

2 Modes and signatures

Modes Let $f : [0, 1] \rightarrow \mathbb{R}$ be an arbitrary function. In order to define the number of *modes* (local maxima) of f , consider a finite partition $P = \{t_0, \dots, t_{|P|}\}$ of $[0, 1]$ such that $0 = t_0 < t_1 < \dots < t_{|P|-1} < t_{|P|} = 1$. For each $0 < i < |P|$ let

$$\mathcal{M}(f, P, i) = \begin{cases} 1 & \text{if } \max(f(t_{i-1}), f(t_{i+1})) < f(t_i) \\ 0 & \text{else.} \end{cases}$$

Define the number of modes of f with respect to P and the total number of modes of f by

$$\mathcal{M}(f, P) = \sum_{i=1}^{|P|-1} \mathcal{M}(f, P, i) \quad \text{and} \quad \mathcal{M}(f) = \sup_P \mathcal{M}(f, P),$$

respectively. It is easy to see that if f is constant, then $\mathcal{M}(f) = 0$ and if f is a Morse function in the classical sense (i.e., a smooth function with only nondegenerate critical points), then $\mathcal{M}(f)$ equals the (possibly infinite) number of local maxima of f on the open interval $(0, 1)$. Notice that different from Morse theory, though, we are not concerned with critical values or critical points of functions; $\mathcal{M}(f)$ merely counts the *number* of modes, without referring to their individual positions or values.

Metric signatures We denote by $\mathcal{L}^\infty$ the linear space of Lebesgue-measurable essentially bounded functions on $[0, 1]$. Notice that we do not regard $\mathcal{L}^\infty$ as a space of equivalence classes of functions. Throughout this article we work with functions in some (to be specified) set $X \subset \mathcal{L}^\infty$. For example, X may consist of functions of bounded variation or piecewise polynomial functions. We do not a priori require X to be a linear space. By (X, d) we denote X together with some metric, but we do not require (X, d) to be a complete metric space. Additionally, we allow that d attains the value ∞ . Particular choices of (X, d) will be specified below.

Definition 1 (Metric signatures) Let X_k denote the subset of X with at most k modes, i.e., $X_k := \{f \in X : \mathcal{M}(f) \leq k\}$. Define the k th *metric signature* of $f \in X$ as

$$s_k(f) := d(f, X_k) = \inf_{g \in X_k} d(f, g) \quad \text{for } k \in \mathbb{N}_0,$$

i.e., the distance of f to the set of functions with at most k modes. □

Clearly, $X_k \subseteq X_{k+1}$ are nested models; hence, the sequence $(s_k(f))_{k \in \mathbb{N}}$ is monotonically decreasing, and $s_k(f)$ measures the minimal distance by which f needs to be moved (with respect to the metric d) in order to remove all but its k most significant modes. What is considered significant and what is not, however, heavily depends on the choice of metric. In any case, so far we have not excluded pathologies, i.e., situations where $\mathcal{M}(f) > k$ but $s_k(f) = d(f, X_k) = 0$. Hence:

Definition 2 (Descriptive metric) (X, d) is called *descriptive* if $\mathcal{M}(f) > k$ implies that $s_k(f) > 0$ for every $f \in X$ and all $k \in \mathbb{N}_0$. □

Stability Regardless of the concrete choice of metric, notice that distance to (arbitrary) sets in metric spaces is 1-Lipschitz; therefore stability essentially comes for free:

Lemma 1 (Stability of signatures) *Let $f, g \in X$. Then*

$$|s_k(f) - s_k(g)| \leq d(f, g)$$

for all $k \in \mathbb{N}_0$. □

Stability implies that a small perturbation of f results in a small perturbation of the signatures $s_k(f)$.

3 Persistence signatures and Kolmogorov signatures

In our setting a “good” metric is one that leads to signatures that clearly separate significant modes (with respect to a given noise model) from insignificant ones. We investigate two choices.

Persistence signatures One possible choice of metric is the one induced by the sup norm, i.e., $d_\infty(f, g) = \sup_x |f(x) - g(x)|$, which leads to signatures that have an interpretation in the context of persistent homology, as we show below.

Lemma 2 *(X, d_∞) is descriptive for every $X \subset \mathcal{L}^\infty$.* □

PROOF Being descriptive is equivalent to X_k being closed in X for all k . Suppose that there exists $k \in \mathbb{N}_0$ such that X_k is not closed, i.e., there exist $f \in X \setminus X_k$ and a sequence (f_n) in X_k with $d_\infty(f_n, f) \rightarrow 0$. Since $f \notin X_k$, there exists a partition $P = \{t_0, \dots, t_{|P|}\}$ of $[0, 1]$ and some index set I with $k < |I| < |P|$ such that $\max(f(t_{i-1}), f(t_{i+1})) < f(t_i)$ for all $i \in I$. Since $d_\infty(f_n, f) \rightarrow 0$, there exists $N \in \mathbb{N}$ such that $\max(f_n(t_{i-1}), f_n(t_{i+1})) < f_n(t_i)$ for all $n \geq N$ and all $i \in I$. Contradiction. ■

The following lemma makes precise the relation between topological persistence and our notion of metric signatures for the sup norm.

Lemma 3 *Let X be a space of tame functions, i.e., $H_*(f^{-1}(-\infty, t])$ has finite rank for all $f \in X$ and all $t \in \mathbb{R}$, and every $f \in X$ has a finite number of homologically critical values. Order the finite persistence values (counted with multiplicity) of some $f \in X$ according to their persistence, from highest to lowest, yielding a persistence sequence $(p_k(f))_{k \geq 1}$. Using d_∞ yields $p_k(f) = 2s_{k-1}(f)$ for all $k \geq 1$.* □

PROOF Let $k \geq 1$. We first claim that $p_k(f) \leq 2s_{k-1}(f)$. Let (f_n) be a sequence in X_{k-1} with $d_\infty(f_n, f) \leq s_{k-1}(f) + \frac{1}{n}$. Notice that $p_k(g) = 0$ for all $g \in X_{k-1} \subset X$. By the stability theorem for persistence diagrams [10], one has $|p_k(g) - p_k(f)| \leq 2d_\infty(g, f)$ for all $f, g \in X$. Together these facts imply that

$$p_k(f) = |p_k(f) - p_k(f_n)| \leq 2d_\infty(f, f_n) \leq 2s_{k-1}(f) + \frac{2}{n},$$

which proves the first claim.

To see that $p_k(f) \geq 2s_{k-1}(f)$, observe that the bound provided by the stability theorem is tight in dimensions less or equal to 2, see [2]. Indeed, if f is tame, then by moving f by at most δ in the sup norm, it is possible to remove all its persistence pairs with persistence less or equal to 2δ without increasing the number of remaining persistence pairs. Hence there exists a function $g \in X_{k-1}$ with $d_\infty(g, f) \leq \frac{1}{2}p_k(f)$, which implies that $s_{k-1}(f) = d_\infty(f, X_{k-1}) \leq d_\infty(f, g) \leq \frac{1}{2}p_k(f)$. ■

Kolmogorov signatures For reasons that will become evident in the next section, we propose an alternative to persistence signatures, which we call *Kolmogorov signatures*. Let $\mathcal{L}^1$ denote the space of Lebesgue-integrable functions on $[0, 1]$. Due to compactness of $[0, 1]$, we have that $\mathcal{L}^\infty \subset \mathcal{L}^1$. The *Kolmogorov distance*, d_K , is defined as follows. Let $f, g \in \mathcal{L}^\infty$, and let F, G denote the respective antiderivatives, where, as a convention, we require that $F(0) = G(0) = 0$. Define

$$d_K(f, g) := d_\infty(F, G) .$$

Notice that d_K does not induce a metric on arbitrary subsets $X \subset \mathcal{L}^\infty$ since if $f = g$ almost everywhere (a.e.), then $d_K(f, g) = 0$. Therefore, we work with a unique representative in each equivalence class of a.e. identical functions by requiring that

$$X \subset \mathcal{L} := \left\{ f \in \mathcal{L}^\infty : f(t) = \lim_{\epsilon \rightarrow 0} \inf_{0 < \delta < \epsilon} \frac{1}{t^+(\delta) - t^-(\delta)} \int_{t^-(\delta)}^{t^+(\delta)} f(s) ds \text{ for all } t \in [0, 1] \right\} , \quad (8)$$

where $t^-(\delta) = \max(0, t - \delta)$ and $t^+(\delta) = \min(1, t + \delta)$.

There indeed exists a (unique) representative in $\mathcal{L}$ for every equivalence class of a.e. identical functions in $\mathcal{L}^\infty$, since the right hand side of (8) exists (and is finite) for all $t \in [0, 1]$ and all $f \in \mathcal{L}^\infty$, and since Lebesgue's differentiation theorem asserts that every $f \in \mathcal{L}^1$ satisfies

$$f(t) = \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} f(s) ds \quad \text{a.e. on } (0, 1) .$$

We thus obtain a projection operator $\mathcal{P} : \mathcal{L}^\infty \rightarrow \mathcal{L} \subset \mathcal{L}^\infty$. Notice, however, that $\mathcal{L}$ is not a linear space, since $f \in \mathcal{L}$ does not necessarily imply that $-f \in \mathcal{L}$. Nonetheless, we may of course choose linear subspaces $X \subset \mathcal{L}$ for specific applications.

The following lemma further motivates our choice of $\mathcal{L}$.

Lemma 4 *For any class $[f]$ of a.e. identical functions in $\mathcal{L}^\infty$, its unique representative $\mathcal{P}(f) \in \mathcal{L}$ minimizes the number of modes within that class.* $\square$

PROOF Let $f \in \mathcal{L}^\infty$ with representative $\tilde{f} := \mathcal{P}(f) \in \mathcal{L}$. We show that $\mathcal{M}(\tilde{f}) \leq \mathcal{M}[f]$. Consider any finite partition P of $[0, 1]$ and assume that t_i counts a mode of $\tilde{f}$, i.e., $\tilde{f}(t_i) - \max(\tilde{f}(t_{i-1}), \tilde{f}(t_{i+1})) > \epsilon$ for some $\epsilon > 0$. Consider any open neighborhood U_i of t_i . Since

$$\tilde{f}(t_i) = \lim_{\epsilon \rightarrow 0} \inf_{0 < \delta < \epsilon} \frac{1}{t_i^+(\delta) - t_i^-(\delta)} \int_{t_i^-(\delta)}^{t_i^+(\delta)} f(s) ds ,$$

there must be some $t \in U_i$ with $f(t) \geq \tilde{f}(t_i) - \epsilon/2$. Since U_i can be chosen arbitrarily small, there exists t'_i arbitrarily close to t_i such that $f(t'_i) \geq \tilde{f}(t_i) - \epsilon/2$. By the same argument, there exist t'_{i-1} and t'_{i+1} arbitrarily close to t_{i-1} and t_{i+1} , respectively, such that $f(t'_{i-1}) \leq \tilde{f}(t_{i-1}) + \epsilon/2$ and $f(t'_{i+1}) \leq \tilde{f}(t_{i+1}) + \epsilon/2$. By our choice of ϵ this implies $f(t'_i) > \max(f(t'_{i-1}), f(t'_{i+1}))$. Continuing this way yields a partition P' with $\mathcal{M}(f, P') \geq \mathcal{M}(\tilde{f}, P)$. $\blacksquare$

Lemma 5 *(X, d_K) is descriptive for every $X \subset \mathcal{L}$.* $\square$

PROOF We show that $X \setminus X_k$ is open wrt. the Kolmogorov metric. Let $f \in X \setminus X_k$. Then there exists a finite partition $P = \{t_0, \dots, t_{|P|}\}$ of $[0, 1]$ and some index set I with $k < |I| < |P|$ such

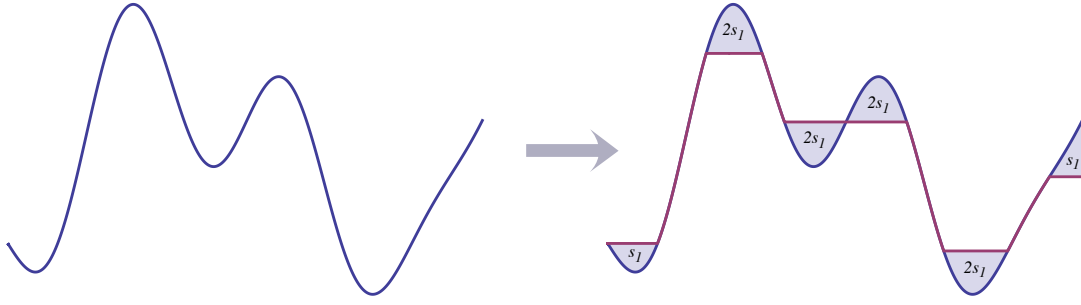


Figure 2: A function with exactly two modes (left) and its closest function with exactly one mode w.r.t. the Kolmogorov norm (right, in purple). Notice that the attendant Kolmogorov signature, s_1 , for removing the smallest mode of f , can be read off from the light-blue areas. The purple function is computed using *taut strings* (see Section 5).

that $f(t_i) - \max(f(t_{i-1}), f(t_{i+1})) > \epsilon$ for all $i \in I$ and some small enough $\epsilon > 0$. Without loss of generality, we assume that $t_{i-1} > 0$ and $t_{i+1} < 1$ for all $i \in I$. Let $\delta > 0$ be small enough such that for all $i \in I$ the intervals $[t_{i-1} - \delta, t_{i-1} + \delta]$, $[t_i - \delta, t_i + \delta]$, and $[t_{i+1} - \delta, t_{i+1} + \delta]$ are contained in $[0, 1]$ and are mutually disjoint. Additionally, for all $i \in I$ and all $j \in \{-1, 0, 1\}$ let $\delta_{i,j} \leq \delta$ be such that

$$\left| f(t_{i+j}) - \frac{1}{2\delta_{i,j}} \int_{t_{i+j}-\delta_{i,j}}^{t_{i+j}+\delta_{i,j}} f(s) ds \right| < \frac{\epsilon}{4}.$$

Let $\delta' = \min_{i \in I, j \in \{-1, 0, 1\}} \delta_{i,j}$. Let $g \in X$ with $d_K(f, g) < \frac{1}{4}\epsilon\delta'$. Then

$$\left| \int_a^b (f(s) - g(s)) ds \right| < \frac{1}{2}\epsilon\delta'$$

for all $0 \leq a < b \leq 1$. Hence,

$$\left| f(t_{i+j}) - \frac{1}{2\delta_{i,j}} \int_{t_{i+j}-\delta_{i,j}}^{t_{i+j}+\delta_{i,j}} g(s) ds \right| < \frac{\epsilon}{4} + \frac{\epsilon\delta'}{4\delta_{i,j}} \leq \frac{\epsilon}{2}$$

for all $i \in I$ and all $j \in \{-1, 0, 1\}$. Therefore, there exists $t'_i \in [t_i - \delta_{i,0}, t_i + \delta_{i,0}]$ with $g(t'_i) > f(t_i) - \frac{\epsilon}{2}$. Likewise, there exist $t'_{i\pm 1} \in [t_{i\pm 1} - \delta_{i\pm 1, \pm 1}, t_{i\pm 1} + \delta_{i\pm 1, \pm 1}]$ with $g(t'_{i\pm 1}) < f(t_{i\pm 1}) + \frac{\epsilon}{2}$. Thus there exists a partition P' of $[0, 1]$ with $\mathcal{M}[g, P'] > k$, i.e., $g \in X \setminus X_k$. Since ϵ and δ' only depend on f and since g was chosen arbitrarily in the open Kolmogorov-ball of radius $\frac{1}{4}\epsilon\delta'$ around f , this ball is contained in $X \setminus X_k$. ■

Figure 2 offers a visualization for a function with two modes and its closest function with a single mode with respect to the Kolmogorov norm. Before elaborating on how to compute Kolmogorov signatures, though, we examine their statistical properties.

4 Statistical perspective

Throughout this section we assume that the noise (ϵ_i) in Model (2) is independently distributed with mean zero such that the moment condition (3) is satisfied.

4.1 Thresholding Kolmogorov signatures

In this subsection we prove an exponential deviation inequality for the empirical Kolmogorov signatures (Theorem 1), which allows us to construct uniform confidence bands for the unknown signatures $(s_j(f))_{j \in \mathbb{N}_0}$. More precisely, we provide a data dependent sequence of intervals $(I^{(n)}(\alpha)_j)_{j \in \mathbb{N}_0}$ that covers the (unknown) signatures with probability at least $1 - \alpha$.

Let $f \in X_k \subset X \subset \mathcal{L}$, with $\mathcal{L}$ defined in (8), have (an unknown number of) exactly k modes. As stressed in the introduction, we do not aim at estimating the regression function f itself but rather at inferring directly the sequence of signatures $s_j(f)$ together with the number of modes k in such a way that estimates for these quantities can be provided at a prespecified error rate. This can be achieved by properly thresholding the sequence of empirical signatures.

In our analysis we consider equidistant sampling points i/n and piecewise constant functions $f^{(n)} : [0, 1] \rightarrow \mathbb{R}$ defined as

$$f^{(n)}(t) = \sum_{i=0}^{n-1} \mathbf{1}_{\left[\frac{i}{n}, \frac{i+1}{n}\right)}(t) f\left(\frac{i}{n}\right).$$

We define $X_j^{(n)}$ as the corresponding set of piecewise constant functions with at most j modes, and we call $s_j(f^{(n)}) = \text{dist}(f^{(n)}, X_j^{(n)})$ the *quantized signature* of f .² Analogously, for the observation vector $Y = (Y_1, \dots, Y_n)$ we define the piecewise constant function

$$Y^{(n)}(t) = \sum_{i=1}^n \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right)}(t) Y_i.$$

In the following, we call $s_j(Y^{(n)})$ the *empirical signatures*.

Function spaces In principle, the results of this subsection hold for any function space $X \subset \mathcal{L}$ as long as one can control the distance $d_K(f, f^{(n)})$ between f and the quantized function $f^{(n)}$. Accordingly, all subsequent results are formulated for the quantized signatures $s_j(f^{(n)})$. From those, the corresponding statements concerning $s_j(f)$ can be obtained along the following reasoning. Consider the (deterministic) approximation error between $f^{(n)}$ and f in terms of the Kolmogorov metric

$$d_K(f, f^{(n)}) = \sup_{s \in [0, 1]} \left| \int_0^s f(t) - f^{(n)}(t) dt \right|. \quad (9)$$

Then, due to Lemma 1 and the triangle inequality, it follows that

$$\max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f)| \leq \max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f^{(n)})| + d_K(f, f^{(n)}).$$

Therefore, if $d_K(f, f^{(n)})$ is known, then the subsequent estimates on $|s_j(Y^{(n)}) - s_j(f^{(n)})|$ can readily be modified to obtain estimates on $|s_j(Y^{(n)}) - s_j(f)|$. E.g., if f Hölder continuous, i.e.,

$$|f(x) - f(y)| \leq C |x - y|^\gamma \quad \forall (x, y) \in [0, 1], \quad \gamma > 0,$$

²Notice that, strictly speaking, the spaces $X_j^{(n)}$ are not contained in $\mathcal{L}$ defined in (8) since this would require to set $f^{(n)}(\frac{i}{n}) = \frac{1}{2} (f^{(n)}(\frac{2i-1}{2n}) + f^{(n)}(\frac{2i+1}{2n}))$. We avoid this technicality by working with $X_j^{(n)}$ as defined above since this does not affect the number of modes.

then

$$d_K(f, f^{(n)}) \leq \frac{C}{\gamma + 1} n^{-\gamma}, \quad (10)$$

so that the approximation error is of order $n^{-\gamma}$. Hence, due to Lemma 1,

$$\max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f)| \leq \max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f^{(n)})| + O(n^{-\gamma}),$$

and all subsequent estimates and results can be modified accordingly.

Statistical inference of signatures and modes without a priori information We return to our initial goal of providing tools for statistical inference on the signatures and modes. We start with investigating how well the empirical signatures $s_j(Y^{(n)})$ estimate the quantized signatures $s_j(f^{(n)})$. To this end, we control $d_K(f^{(n)}, Y^{(n)})$ by the following exponential deviation bound, which is a direct consequence of [27, Theorem B.2].

Theorem 1 *Assume the moment condition in (3). Then, for any $\delta > 0$ and any $f \in X$, one has*

$$\mathbb{P}\left(\max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f^{(n)})| \geq \delta\right) \leq 2 \exp\left(-\frac{\delta^2 n}{2v + 2\kappa\delta}\right). \quad \square$$

PROOF By stability of metric signatures (Lemma 1), we have that

$$\mathbb{P}\left(\max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f^{(n)})| \geq \delta\right) \leq \mathbb{P}\left(d_K(f^{(n)}, Y^{(n)}) \geq \delta\right).$$

Let $S_k = \sum_{i=1}^k \epsilon_i$ and observe that $d_K(Y^{(n)}, f^{(n)}) = \max_k |S_k|/n$. From [27, Theorem B.2] we obtain

$$\mathbb{P}\left(d_K(Y^{(n)}, f^{(n)}) \geq \delta\right) = \mathbb{P}\left(\max_k |S_k| \geq \delta n\right) \leq 2 \exp\left(-\frac{\delta^2 n}{2v + 2\kappa\delta}\right). \quad \blacksquare$$

This results shows that the empirical signatures $s_j(Y^{(n)})$ are close to the quantized signatures $s_j(f^{(n)})$ with high probability *simultaneously* for all $j \in \mathbb{N}_0$.

Remark 1 (Sharpness of bound) Figure 3 offers two examples of how the signatures of $Y^{(n)}$ deviate from those of $f^{(n)}$. Notice that in these examples, the signatures of $f^{(n)}$ are almost indistinguishable from the highest signatures of $Y^{(n)}$ —indeed, their difference is *less* than what is predicted by Theorem 1. The reason is that, while the bound in Theorem 1 is sharp in general (since stability of metric signatures provides a sharp bound in general), it may be arbitrarily suboptimal for concrete examples, i.e., if $|s_k(f) - s_k(Y)|$ is small while $d_K(f, Y)$ is large. $\square$

A useful application of Theorem 1 is that for a given probability α , we can construct a non-asymptotic and honest (uniform) confidence region covering the signatures $s_j(f^{(n)})$ with probability at least $1 - \alpha$, as shown in the following theorem.

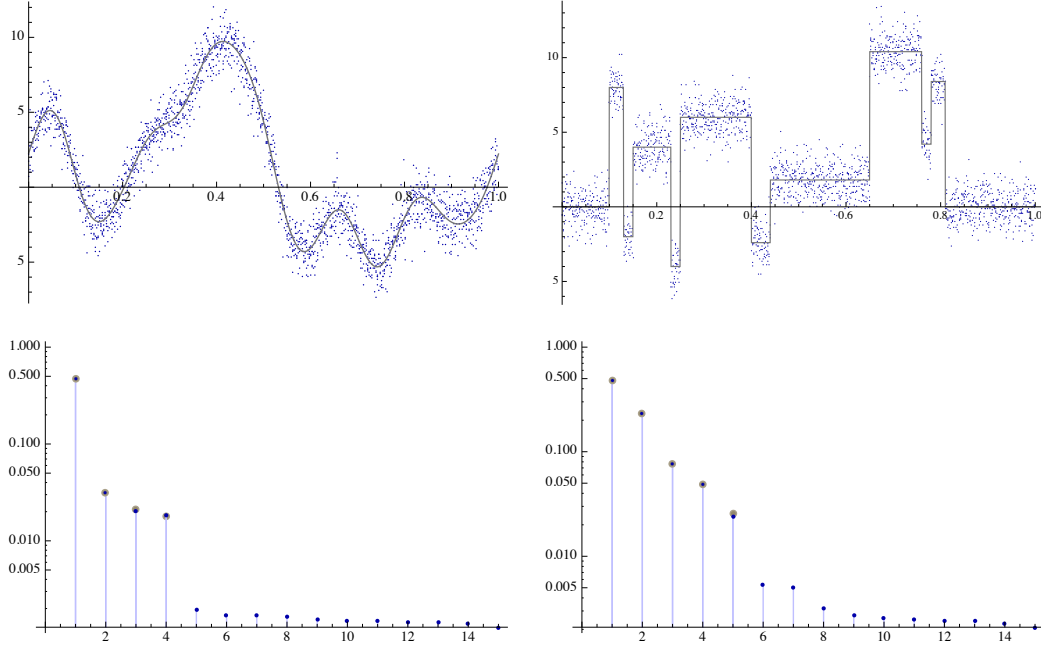


Figure 3: Noisy samples of signals (top) and signatures of both original signal and sample (bottom, log-scale). Left: Function generated by random sampling and smoothing. Right: signal bumps [15]. Sampling noise normally distributed with standard deviation $\sigma = 1$. Notice that the largest signatures of signal and sample are very close (almost indistinguishable), and that there is a clear gap between the smallest signature of the signal (left: $k = 4$, right: $k = 5$) and the next signature of the noisy sample.

Theorem 2 Fix some $\alpha \in (0, 1)$ and let

$$\tau_n(\alpha) := \frac{1}{n} \left(\sqrt{\log(\alpha/2) (\log(\alpha/2) \kappa^2 - 2nv)} - \kappa \log(\alpha/2) \right).$$

Assume the regression model (2) and the moment condition in (3). Then

$$\inf_{f \in X} \mathbb{P} \left(s_j(f^{(n)}) \in \left[(s_j(Y^{(n)}) - \tau_n(\alpha))_+, s_j(Y^{(n)}) + \tau_n(\alpha) \right] \text{ for all } j \in \mathbb{N}_0 \right) \geq 1 - \alpha,$$

where $(x)_+ = \max(0, x)$. □

PROOF From Theorem 1 we obtain

$$\begin{aligned} \mathbb{P} \left(|s_j(Y^{(n)}) - s_j(f^{(n)})| \leq \tau_n(\alpha) \text{ for all } j \in \mathbb{N}_0 \right) &\geq 1 - 2 \exp \left(-\frac{\tau_n(\alpha)^2 n}{2v + 2\kappa \tau_n(\alpha)} \right) \\ &= 1 - \alpha. \end{aligned}$$

Since $s_j(f^{(n)}) \geq 0$ for all $j \in \mathbb{N}_0$, this completes the proof. ■

Note that $\tau_n(\alpha)$ is a quantity that only depends on the values n, κ, v , and the confidence level α . Here we assume for simplicity that κ and v are known—and while in practice this might not be the case, these numbers can be estimated from the data, e.g., in the case of a normal distribution, such

an estimate boils down to estimating the variance σ^2 . Fixing α , we obtain a (random) sequence of intervals

$$\left[\left(s_j(Y^{(n)}) - \tau_n(\alpha) \right)_+, s_j(Y^{(n)}) + \tau_n(\alpha) \right],$$

which, according to Theorem 2, cover the sequence of true quantized signatures $s_j(f^{(n)})$ with confidence level $1 - \alpha$. For smaller values of α , i.e., for larger confidence, these intervals become wider. Notice that for a fixed error $\alpha \in (0, 1)$, the interval lengths $2\tau_n(\alpha)$ behave like $1/\sqrt{n}$ as $n \rightarrow \infty$.

Theorem 1 shows that $s_j(Y^{(n)})$ approximates $s_j(f^{(n)})$ well in the sup norm. However, the number of estimated signatures greater than zero might still be large. Consequently, $s(Y^{(n)})$ does not directly indicate which signatures are significantly larger than zero and hence will be of limited use for estimating the number of modes of f . Nonetheless, such an estimate can readily be obtained by thresholding the empirical signatures. Define

$$k_\epsilon(Y^{(n)}) = \max\{j \in \mathbb{N}_0 : s_{j-1}(Y^{(n)}) \geq \epsilon\}, \quad (11)$$

where, as a convention, we define $s_{-1}(Y^{(n)}) = \infty$.

The threshold parameter $\tau_n(\alpha)$ has an immediate statistical interpretation: It controls the probability of overestimating the number of modes for any function $f \in X$.

Theorem 3 *Let $f \in X$, assume the regression model (2) and the moment condition (3), let $\alpha \in (0, 1)$, and let $k \in \mathbb{N}_0$ be such that $f^{(n)} \in X_k$. Then*

$$\mathbb{P}\left(k_{\tau_n(\alpha)}(Y^{(n)}) > k\right) \leq \alpha. \quad \square$$

PROOF First, observe from the definition of $k_{\tau_n(\alpha)}(Y^{(n)})$ in (11) that

$$\mathbb{P}\left(k_{\tau_n(\alpha)}(Y^{(n)}) > k\right) = \mathbb{P}\left(s_k(Y^{(n)}) \geq \tau_n(\alpha)\right).$$

Notice that $f^{(n)} \in X_k$ implies that $s_k(f^{(n)}) = 0$. Therefore, for $f^{(n)} \in X_k$, Theorem 1 and the definition of $\tau_n(\alpha)$ imply (similar to the proof of Theorem 2) that

$$\mathbb{P}\left(s_k(Y^{(n)}) \geq \tau_n(\alpha)\right) \leq \alpha. \quad \blacksquare$$

Hence, whatever the number of modes of $f^{(n)}$ might be, the thresholding index $k_{\tau_n(\alpha)}(Y^{(n)})$ overestimates this number with probability less or equal to α . Notice that the thresholding parameter $\tau_n(\alpha)$ is independent of the number and magnitude of the modes of f , so in that sense, this result is universal.

As mentioned in the introduction, obtaining a universal result in the other direction, i.e., controlling the probability of *underestimating* the number of modes, is a more delicate task since modes can become arbitrarily small. Recalling the definition of $k_\epsilon(f)$ as in (11), we find:

Theorem 4 *Let $f \in X$, assume the regression model (2) and the moment condition (3), let $\alpha \in (0, 1)$, and let $k \in \mathbb{N}_0$ be such that $f^{(n)} \in X_k$. Then*

$$\mathbb{P}\left(k_{\tau_n(\alpha)}(Y^{(n)}) < k_{2\tau_n(\alpha)}(f^{(n)})\right) \leq \alpha. \quad (12)$$

□

PROOF Let $f^{(n)} \in X_k$ and let l denote the largest integer such that $s_{l-1}(f^{(n)}) \geq 2\tau_n(\alpha)$, i.e., $l = k_{2\tau_n(\alpha)}(f^{(n)})$. If $l = 0$, then (12) is trivially satisfied, since $k_{\tau_n(\alpha)}(Y^{(n)}) \geq 0$. So suppose that $l > 0$. Then

$$\begin{aligned} \mathbb{P}\left(k_{\tau_n(\alpha)}(Y^{(n)}) < l\right) &= \mathbb{P}\left(s_{l-1}(Y^{(n)}) < \tau_n(\alpha)\right) \\ &\leq \mathbb{P}\left(s_{l-1}(f^{(n)}) - s_{l-1}(Y^{(n)}) > \tau_n(\alpha)\right) \\ &\leq \mathbb{P}\left(|s_{l-1}(f^{(n)}) - s_{l-1}(Y^{(n)})| > \tau_n(\alpha)\right) \\ &\leq \alpha, \end{aligned}$$

where the last inequality follows from Theorem 1 and the definition of $\tau_n(\alpha)$. $\blacksquare$

We have thus expressed the underestimation error of the number of modes as an explicit function of the signature threshold $2\tau_n(\alpha)$. Combining the latter results, we obtain *two sided bounds* for the estimated number of modes. More precisely, for any f and k with $f^{(n)} \in X_k$ and any $\alpha \in (0, 1)$ we have that

$$\mathbb{P}\left(k_{2\tau_n(\alpha/2)}(f^{(n)}) \leq k_{\tau_n(\alpha/2)}(Y^{(n)}) \leq k\right) \geq 1 - \alpha.$$

As mentioned above, for fixed α, κ, ν one has $\tau_n(\alpha) \approx 1/\sqrt{n}$. Therefore there exists a constant C such that asymptotically (for large enough n) by thresholding at $C/\sqrt{n}$, it can be guaranteed at a level α that all signatures above this threshold are detected.

Based on the previous results we now construct confidence intervals for $k_\epsilon(f^{(n)})$, i.e., for the number of modes whose signatures exceed a certain size ϵ .

Corollary 1 Assume the regression model (2), the moment condition (3), let $\epsilon \geq 0$, and let $f^{(n)} \in X_k$. Define

$$l(\alpha, \epsilon) = \begin{cases} \max\{j \in \mathbb{N}_0 : s_j(Y^{(n)}) > \epsilon + \tau_n(\alpha)\} & \text{if } \epsilon < s_0(Y^{(n)}) - \tau_n(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

and

$$u(\alpha, \epsilon) = \begin{cases} \min\{j \in \mathbb{N}_0 : s_j(Y^{(n)}) < \epsilon - \tau_n(\alpha)\} & \text{if } \epsilon > \tau_n(\alpha) \\ \infty & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{P}\left(k_\epsilon(f^{(n)}) \in [l(\alpha, \epsilon), u(\alpha, \epsilon)]\right) \geq 1 - \alpha. \quad \square$$

PROOF Suppose, for the moment, that

$$d_K(Y^{(n)}, f^{(n)}) \leq \tau_n(\alpha). \quad (13)$$

Since $s_j(f^{(n)}) \geq \epsilon$ for all $j < k_\epsilon(f^{(n)})$, stability of metric signatures implies that $s_j(Y^{(n)}) \geq \epsilon - \tau_n(\alpha)$ for all $j < k_\epsilon(f^{(n)})$. Hence, by the definition of $u(\alpha, \epsilon)$, we have $u(\alpha, \epsilon) \geq k_\epsilon(f^{(n)})$.

Further, while still assuming (13), $s_j(Y^{(n)}) > \epsilon + \tau_n(\alpha)$ implies that $s_j(f^{(n)}) > \epsilon$. Hence, by the definition of $l(\alpha, \epsilon)$, we find that $s_{l(\alpha, \epsilon)}(f^{(n)}) > \epsilon$. This in turn implies $k_\epsilon(f^{(n)}) \geq l(\alpha, \epsilon)$. Therefore, we have so far shown that (13) implies that

$$l(\alpha, \epsilon) \leq k_\epsilon(f^{(n)}) \leq u(\alpha, \epsilon).$$

Since (13) holds with probability $\geq 1 - \alpha$ (see proof of Theorem 1), this proves the assertion. $\blacksquare$

Note that the upper bound for k_ϵ jumps to ∞ if $\epsilon \leq \tau_n(\alpha)$. This reflects the fact that meaningful upper bounds cannot be provided for signatures whose size is of the order of the noise level.

Remark 2 (Distribution of signatures) Assume the setting of Theorem 1 and suppose that X_k is scaling invariant for all $k \in \mathbb{N}_0$, i.e., $\{\lambda g : g \in X_k\} = X_k$ for all $0 < \lambda \in \mathbb{R}$. Assume for simplicity that $f \equiv 0$, the general case still being unknown. Then, for any $k \in \mathbb{N}_0$, we have that

$$\begin{aligned} \sqrt{n} \left(s_k(Y^{(n)}) - s_k(f^{(n)}) \right) &= \sqrt{n} \inf_{g \in X_k} \sup_{s \in [0,1]} \left| \int_0^s (\epsilon^{(n)}(t) - g(t)) dt \right| \\ &= \inf_{g \in X_k} \sup_{s \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \epsilon_i - \int_0^s \sqrt{n} g(t) dt \right| \\ &= \inf_{g \in X_k} \sup_{s \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \epsilon_i - \int_0^s g(t) dt \right|, \end{aligned}$$

where the last equality follows from the scaling invariance of X_k . Noting that

$$\frac{1}{\sqrt{n}} \max_{m=1, \dots, n} \left| \sum_{i=1}^m \epsilon_i \right| \xrightarrow{\mathcal{D}} \sup_{0 \leq x \leq 1} B(x),$$

where B denotes a standard Brownian motion on $[0, 1]$ and using that $f \equiv 0$, it follows that

$$\sqrt{n} \left(s(Y^{(n)}) \right) \xrightarrow{\mathcal{D}} s(B'),$$

where B' denotes the derivative of a standard Brownian Motion on $[0, 1]$ in a weak sense. This follows from the continuity of the functional s w.r.t. the Kolmogorov norm. $\square$

Remark 3 (Gaussian observation) If the noise ϵ in (2) is Gaussian with mean zero and variance σ^2 , then Theorem 1 can be sharpened, due to a refined large deviation result for Gaussian observations (see, e.g., [4]):

$$\mathbb{P} \left(d_K(f^{(n)}, Y^{(n)}) \geq \delta \right) \leq 2 \exp \left(-\frac{\delta^2 n}{2\sigma^2} \right). \quad (14)$$

Hence, in the Gaussian case, all results of Section 4 remain true if $\tau_n(\alpha)$ is replaced by the simpler (and slightly sharper) threshold

$$\tilde{\tau}_n(\alpha) = \sqrt{-2\sigma^2/n \log(\alpha/2)}. \quad \square$$

Obtaining the correct number of modes using a priori information Notice that so far we have not made any a priori assumption about $f^{(n)}$. If, however, $f^{(n)} \in X_k$, and if we impose prior information on the smallest strictly positive signature $s_{k-1}(f^{(n)})$, then we obtain an explicit bound for the probability that the number of modes is estimated correctly.

Theorem 5 Assume the regression model (2) and the moment condition (3). Let $f^{(n)} \in X_k$ be such that $s_{k-1}(f^{(n)}) \geq \epsilon$. Then

$$\mathbb{P} \left(k_{\epsilon/2}(Y^{(n)}) = k \right) \geq 1 - 2 \exp \left(-\frac{\epsilon^2 n}{8v + 4k\epsilon} \right). \quad (15)$$

$\square$

PROOF First suppose that $k > 0$. Notice that by (11) we have that $k_{\epsilon/2}(Y) = k$ iff $s_{k-1}(Y^{(n)}) \geq \frac{\epsilon}{2}$ and $s_k(Y^{(n)}) < \frac{\epsilon}{2}$. Furthermore, by assumption we have that $s_{k-1}(f^{(n)}) \geq \epsilon$ and $s_k(f^{(n)}) = 0$. Therefore, $k_{\epsilon/2}(Y) \neq k$ implies that

$$|s_{k-1}(f^{(n)}) - s_{k-1}(Y^{(n)})| \geq \frac{\epsilon}{2} \quad \text{or} \quad |s_k(f^{(n)}) - s_k(Y^{(n)})| \geq \frac{\epsilon}{2}.$$

For $k = 0$, by a similar argument, we have that $k_{\epsilon/2}(Y^{(n)}) \neq 0$ implies that $|s_k(f^{(n)}) - s_k(Y^{(n)})| \geq \frac{\epsilon}{2}$. Thus, for all $k \geq 0$, Theorem 1 implies that

$$\mathbb{P}(k_{\epsilon/2}(Y^{(n)}) \neq k) \leq \mathbb{P}\left(\max_{j \in \mathbb{N}_0} |s_j(Y^{(n)}) - s_j(f^{(n)})| \geq \frac{\epsilon}{2}\right) \leq 2 \exp\left(-\frac{\epsilon^2 n}{8v + 4\kappa\epsilon}\right). \quad \blacksquare$$

We stress that the bound in Theorem 5 is remarkably simple, as it depends on the signatures $s_j(f^{(n)})$, $j \in \{0, \dots, k-1\}$, only through $s_{k-1}(f^{(n)})$, which in a sense represents the signature that is hardest to detect. Notice furthermore that the bound in Theorem 5 does not depend on the (unknown) number of signatures k .

Limitations of Kolmogorov signatures Kolmogorov signatures are by no means suitable for all kinds of signals. Indeed, as might be expected intuitively, Kolmogorov signatures are not well suited for sparse signals that have high peaks with small support (the *needle in a haystack problem*). In order to illustrate this effect, consider signals of the following kind:

$$f_n(x) = \begin{cases} (1 + \varepsilon) \sqrt{2 \log n} & \text{if } x \in [j/, (j+1)/n], \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

for some $\varepsilon > 0$ and for some $j \in \{0, \dots, n-1\}$ that is *a priori not known*. Note that there exists *no* statistical testing procedure that can asymptotically (as the number of observation $n \rightarrow \infty$) detect signals with intensity as in (16) for $\varepsilon < 0$ with positive detection power, see, e.g., [13]. For $\varepsilon > 0$, sup norm based thresholding is known to achieve the optimal detection boundary [13]. In contrast, Kolmogorov signature based thresholding at $\tau_n(\alpha)$ as described above is not able to detect signals of the type (16) for any $\varepsilon > 0$:

Theorem 6 (Kolmogorov signatures and sparse signals) Let $f_n : [0, 1] \rightarrow \mathbb{R}$ be as in (16), and let $Y_i = f_n(\frac{i}{n}) + \epsilon_i$, where $\epsilon_1, \dots, \epsilon_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Then for any $\alpha \in (0, 1)$ one has

$$\lim_{n \rightarrow \infty} \mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) = 1) = 0,$$

i.e., it is impossible to detect the single mode of f_n when thresholding Kolmogorov signatures at $\tau_n(\alpha)$. $\square$

PROOF We have

$$\begin{aligned} \mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) = 1) &\leq \mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) \geq 1) \\ &= \mathbb{P}(s_0(Y^{(n)}) \geq \tau_n(\alpha)) \\ &\leq \mathbb{P}(d_K(Y^{(n)}, 0) \geq \tau_n(\alpha)), \end{aligned}$$

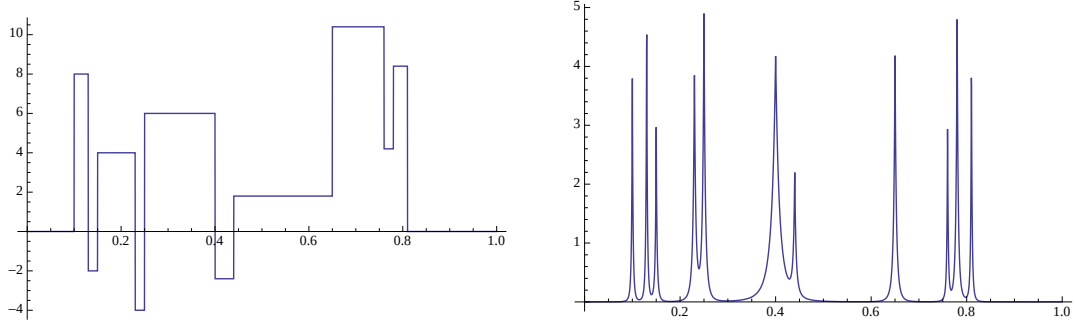


Figure 4: Left: signal *blocks*; Right: signal *bumps* [15].

where $d_K(Y^{(n)}, 0)$ denotes the Kolmogorov distance of the observations $Y^{(n)}$ to the zero function. Let $\mu_n := (1 + \varepsilon) \sqrt{2 \log n}$. Then the last term can be further estimated as

$$\begin{aligned} \mathbb{P}\left(d_K(Y^{(n)}, 0) \geq \tau_n(\alpha)\right) &= \mathbb{P}\left(\frac{1}{n} \max_{m=1, \dots, n} \left| \sum_{i=1}^m \epsilon_i + \mu_n \right| \geq \tau_n(\alpha)\right) \\ &\leq \mathbb{P}\left(\frac{1}{n} \max_{m=1, \dots, n} \left| \sum_{i=1}^m \epsilon_i \right| + \frac{\mu_n}{n} \geq \tau_n(\alpha)\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{m=1, \dots, n} \left| \sum_{i=1}^m \epsilon_i \right| + \frac{\mu_n}{\sqrt{n}} \geq \sqrt{n} \tau_n(\alpha)\right). \end{aligned}$$

The claim now follows from the fact that with $n \rightarrow \infty$ one has $\mu_n / \sqrt{n} \rightarrow 0$, $\sqrt{n} \tau_n(\alpha) \rightarrow \infty$, and

$$\frac{1}{\sqrt{n}} \max_{m=1, \dots, n} \left| \sum_{i=1}^m \epsilon_i \right| \xrightarrow{\mathcal{D}} \sup_{0 \leq x \leq 1} B(x),$$

where B denotes a standard Brownian motion on $[0, 1]$. ■

4.2 Simulations using Kolmogorov signatures

We illustrate the validity of our approach by means of a simulation study for the signals *blocks* and *bumps* [15], which are shown in Fig. 4. Concerning detection of modes, the two signals are of different types as they contain modes of different lengths. For a function f with k modes and observations Y from (2) the theory in the previous Section shows that the number of modes k can be estimated by thresholding of the empirical signatures. This approach clearly relies on the fact that $s_{k-1}(Y^{(n)})$ and $s_k(Y^{(n)})$ can be distinguished with high probability. Here, we investigate this empirically by considering the quantity

$$\Delta(Y) = \frac{s_{k-1}(Y^{(n)})}{s_k(Y^{(n)})}.$$

For our simulation we consider independent Gaussian noise. We note that the bound in Remark 3 is constant for increasing n if the variance is linearly increasing in n . This suggests that the expected value of $\Delta(Y)$ is also constant in this case.

We chose $\sigma = \sqrt{n}/16$ for *blocks* and $\sigma = \sqrt{n}/256$ for *bumps* and computed the average value of $\Delta(Y)$ in 1000 Monte-Carlo simulations. The results in Section 4.2 show that $\Delta(Y)$ is

approximately constant for $n \geq 1024$. Further, for both signals the ratio $\Delta(Y)$ is bounded away from 1, which empirically confirms that the number of modes can be estimated by thresholding.

n	blocks	bumps
256	1.28726	2.08565
1024	1.57086	1.8708
4096	1.52344	1.85699
16384	1.52735	1.84809
65536	1.52647	1.83197

Table 1: Average values of $\Delta(Y)$ for *blocks* and *bumps* as in Fig. 4. The results are obtained from 1000 simulations with independent Gaussian noise with $\sigma = \sqrt{n}/16$ and $\sigma = \sqrt{n}/256$ for *blocks* and *bumps*, respectively. As becomes evident from Fig. 4, the correct number of modes of the signal is $k = 5$ and $k = 11$, respectively.

4.3 Persistence signatures

We contrast the results of the previous sections with what holds true for persistence signatures. Throughout this section, let $s_{j,\infty}$ denote the signatures with respect to the sup norm. For simplicity we restrict our exposition to functions with one *single* mode. More precisely, we consider functions of the type

$$f_n(x) = \begin{cases} \delta_n & \text{if } x \in [1/3, 2/3], \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

with $\delta_n \rightarrow 0$. It is well known that it is possible to detect the single mode of f_n with probability tending to one as $n \rightarrow \infty$ if (on top of $\delta_n \rightarrow 0$) one has

$$\delta_n \sqrt{n} \rightarrow \infty, \quad (18)$$

see, e.g., [7, 31]. From Theorem 5 it follows, using $\epsilon = \delta_n$, that Kolmogorov signatures can correctly detect the single mode of signals in (7) by thresholding signatures at $\delta_n/2$. In contrast, for persistence signatures there exists no thresholding strategy that can detect the single mode with probability one:

Theorem 7 *Let $Y_i = f_n(i/n) + \epsilon_i$, where $\epsilon_1, \dots, \epsilon_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and let $f_n : [0, 1] \rightarrow \mathbb{R}$ be as in (17) with δ_n such that $\delta_n \sqrt{\log n} \rightarrow 0$. Then for an arbitrary sequence (q_n) , with $q_n \in \mathbb{R}$, one has*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(k_{q_n}^\infty(Y) = 1) < 1. \quad \square$$

The proof of Theorem 7 requires some preparation. First, recall that a sequence of random variables $Z_1, \dots, Z_n$ follows a *Gumbel extreme value limit* (GEVL) with sequences a_n and b_n if

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{1 \leq i \leq n} Z_i \leq a_n + b_n x\right) = e^{-e^{-x}}.$$

A sequence of i.i.d. standard normal random variables follows a GEVL with

$$a_n = \sqrt{2 \log n} - (1/2 \log \log n + \log 2 \sqrt{\pi}) / \sqrt{2 \log n}, \quad b_n = 1 / \sqrt{2 \log n}. \quad (19)$$

Another essential ingredient of the proof of Theorem 7 is the following lemma.

Lemma 6 Let $m \in \mathbb{N}$, assume $\epsilon_1, \dots, \epsilon_{2m} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and set

$$\Delta_m = \min_{h \in \mathbb{R}^{2m}: h_1 \leq h_2 \leq \dots \leq h_{2m}} \|\epsilon - h\|_\infty.$$

Then, with a_m and b_m as in (19),

$$\lim_{m \rightarrow \infty} \mathbb{P}(\Delta_m \leq a_m + b_m x) \leq e^{-e^{-x}}.$$

□

PROOF (OF LEMMA 6) Consider a fixed vector $h \in \mathbb{R}^{2m}$ such that $h_1 \leq h_2 \leq \dots \leq h_{2m}$. In particular, $h_j \leq h_m$ for all $j \leq m$ and $h_j \geq h_m$ for all $j \geq m$. Let $M^{(1)} = \max_{i=1, \dots, m} \epsilon_i$ and $M_{(2)} = \min_{i=m+1, \dots, 2m} \epsilon_i$, and observe

$$\|\epsilon - h\|_\infty \geq \max \{M^{(1)} - h_m, h_m - M_{(2)}\}.$$

Hence,

$$\Delta_m \geq \min_{\zeta \in \mathbb{R}} \max \{M^{(1)} - \zeta, \zeta - M_{(2)}\} = \frac{1}{2} (M^{(1)} - M_{(2)}) \stackrel{\mathcal{D}}{=} M^{(1)},$$

where $A \stackrel{\mathcal{D}}{=} B$ means that A and B are equally distributed. This implies that

$$\lim_{m \rightarrow \infty} \mathbb{P}(\Delta_m \leq a_m + b_m x) \leq \lim_{m \rightarrow \infty} \mathbb{P}(M^{(1)} \leq a_m + b_m x) = e^{-e^{-x}}, \quad (20)$$

because $M^{(1)}$ is the maximum of m independent standard normal random variables and follows a GEVL with a_m and b_m . ■

PROOF (OF THEOREM 7) To ease notation, we assume that $n = 6m$ for some $m \in \mathbb{N}$ and hence $m = m(n)$. First, we observe that

$$\mathbb{P}(k_{q_n}^\infty(Y^{(n)}) \geq 1) = \mathbb{P}(s_{0,\infty}(Y^{(n)}) \geq q_n) \quad \text{and} \quad \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) > 1) = \mathbb{P}(s_{1,\infty}(Y^{(n)}) \geq q_n). \quad (21)$$

Since $s_{0,\infty}(Y^{(n)}) \leq d_\infty(f_n, Y) + s_0(f_n)$ (by Lemma 1) and $s_0(f_n) = \delta_n/2$ it holds that

$$\begin{aligned} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) \geq 1) &\leq \mathbb{P}(d_\infty(f_n, Y) \geq q_n - \delta_n/2) \\ &= \mathbb{P}\left(\frac{d_\infty(f_n, Y) - a_n}{b_n} \geq \frac{q_n - a_n}{b_n} - \frac{\delta_n}{2b_n}\right) \\ &= \mathbb{P}\left(\frac{d_\infty(f_n, Y) - a_n}{b_n} \geq \frac{q_n - a_n}{b_n} + o(1)\right), \end{aligned} \quad (22)$$

with a_n and b_n as in (19). Since $d_\infty(f_n, Y) = \max_{i=1, \dots, n} |\epsilon_i|$ it follows that for any $x \in \mathbb{R}$ one has $\mathbb{P}(d_\infty(f_n, Y) \geq x) \leq 2\mathbb{P}(\max_{i=1, \dots, n} \epsilon_i \geq x)$ by symmetry. Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{d_\infty(f_n, Y) - a_n}{b_n} \geq x\right) \leq 2(1 - e^{-e^{-x}}). \quad (23)$$

Further, for $i = 0, \dots, 5$ we define

$$\Delta_i^\pm = \min_{h \in \mathbb{R}^m: h_1 \leq h_2 \leq \dots \leq h_m} \|\pm(Y_{im+1}, \dots, Y_{(i+1)m}) - h\|_\infty.$$

Recall that $s_{1,\infty}(Y^{(n)}) = \inf_{g \in X_1} d_\infty(g, Y)$. Observe that any $g \in X_1$ is either monotonically increasing or decreasing on $[i/6, (i+1)/6]$ for some $0 \leq i \leq 5$. Otherwise g would have two modes, which contradicts $g \in X_1$. For this reason, we find $s_{1,\infty}(Y^{(n)}) \geq \min\{\Delta_0^-, \Delta_0^+, \dots, \Delta_5^-, \Delta_5^+\}$. Note that $\Delta_0^-, \Delta_0^+, \dots, \Delta_5^-, \Delta_5^+$ are identically distributed and independent asymptotically. Therefore,

$$\begin{aligned} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) > 1) &= \lim_{n \rightarrow \infty} \mathbb{P}(s_{1,\infty}(Y^{(n)}) \geq q_n) \geq \lim_{n \rightarrow \infty} \mathbb{P}(\min\{\Delta_0^-, \Delta_0^+, \dots, \Delta_5^-, \Delta_5^+\} \geq q_n) \\ &= \lim_{n \rightarrow \infty} (1 - \mathbb{P}(\Delta_0^- < q_n))^{12}. \end{aligned}$$

In order to prove the assertion, we show that for some $\beta \in (0, 1)$

$$\lim_{n \rightarrow \infty} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) \geq 1) \geq 1 - \beta$$

already implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) > 1) > 0$$

for any sequence $q_n \in \mathbb{R}$. In other words, no thresholding procedure can estimate the number of true modes $k = 1$ with probability tending to one. Combining (22) and (23) shows that $\lim_{n \rightarrow \infty} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) \geq 1) \geq 1 - \beta$ implies

$$q_n \leq a_n + b_n z_\beta + o(b_n)$$

where z_β is defined by $2(1 - \exp(-\exp(-z_\beta))) = \beta$ (it is assumed w.l.o.g. that $\beta < e^{1/6}/2$). We then find from (13) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(k_{q_n}^\infty(Y^{(n)}) > 1) &\geq \lim_{n \rightarrow \infty} (1 - \mathbb{P}(\Delta_0^- < a_n + b_n z_\beta + o(b_n)))^{12} \\ &= \lim_{n \rightarrow \infty} \left(1 - \mathbb{P}\left(\frac{\Delta_0^- - a_m}{b_m} < \frac{a_n - a_m}{b_m} + \frac{b_n}{b_m} z_\beta\right)\right)^{12} \\ &\geq \left(1 - e^{-e^{-z_\beta - \log 6}}\right)^{12}. \end{aligned}$$

Here the last inequality follows from Lemma 6 together with $\frac{b_n}{b_m} \rightarrow 1$ and $\frac{a_n - a_m}{b_m} \rightarrow \log 6$. The proof is then completed by observing that $z_\beta + \log 6 < \infty$, which yields $(1 - e^{-e^{-z_\beta - \log 6}}) > 0$. ■

5 Taut strings

In order to compute Kolmogorov signatures, we require some well known and also some less known results about *taut strings*, see e.g. [11, 26]. We prove a result that is central for our exposition and appears to be interesting in its own right: Taut strings minimize the number of critical points within a certain (quite general) class of functions.

For a given $f \in \mathcal{L}$ with antiderivative F , consider the d_∞ -ball $D_\alpha(F)$ of radius $\alpha \geq 0$ around F . We refer to $D_\alpha(F)$ as the α -tube around F . The *taut string*, denoted by F_α , is the unique function in $D_\alpha(F)$ whose graph, regarded as a curve in $\mathbb{R}^2$, has minimal total curve length, subject to boundary conditions

$$F_\alpha(0) = F(0) \quad \text{and} \quad F_\alpha(1) = F(1).$$

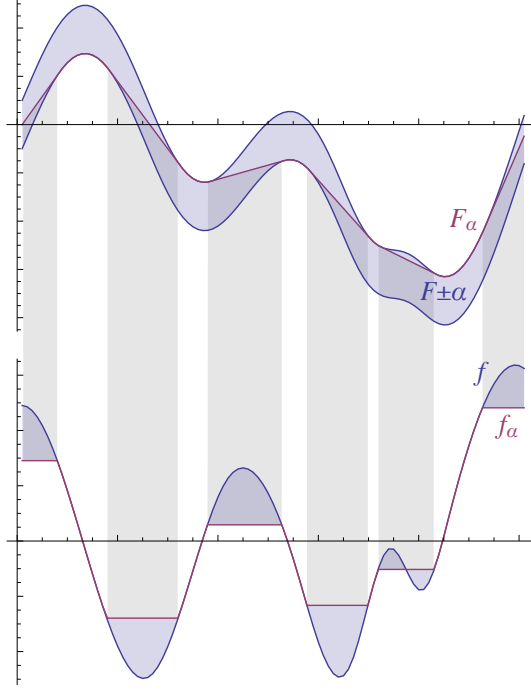


Figure 5: Taut string F_α (purple) in the α -tube around F (top) and its derivative f_α (bottom).

For existence and uniqueness, we refer to [20, 21]. F_α is *Lipschitz continuous* for all $\alpha > 0$ (see [20], proof of Lemma 2); thus its derivative f_α (defined a.e.) is in $\mathcal{L}^\infty$ and we may hence choose $f_\alpha \in \mathcal{L}$.

Therefore, the properties that $F_\alpha \in D_\alpha(F)$ and that the graph of F_α has minimal curve length are equivalent to

$$d_K(f, f_\alpha) \leq \alpha \quad \text{and} \quad \int_0^1 \sqrt{1 + f_\alpha^2(t)} \, dt = \min ,$$

respectively. The aim of this section is to show the following result.

Theorem 8 *For all $f \in \mathcal{L}$ and all $\alpha > 0$, the derivative $f_\alpha \in \mathcal{L}$ of the taut string F_α minimizes the number of modes among all function $g \in \mathcal{L}$ with $d_K(f, g) \leq \alpha$. $\square$*

The proof requires some preparation. Let the top and bottom functions of the α -tube around the antiderivative F of $f \in \mathcal{L}$ be denoted by

$$T_\alpha(t) := F(t) + \alpha \quad \text{and} \quad B_\alpha(t) := F(t) - \alpha ,$$

respectively. Furthermore, let

$$S_{T,\alpha} = \{t \in [0, 1] : F_\alpha(t) = T_\alpha(t)\} \quad \text{and} \quad S_{B,\alpha} = \{t \in [0, 1] : F_\alpha(t) = B_\alpha(t)\}$$

denote the sets where the taut string touches the top (resp. bottom) of the α -tube.

Lemma 7 ([21]) For every $\alpha > 0$, the taut string F_α is the unique function in $D_\alpha(F)$ with $F_\alpha(0) = F(0)$ and $F_\alpha(1) = F(1)$ that is convex on every connected component of $(0, 1) \setminus S_{B,\alpha}$ and concave on every connected component of $(0, 1) \setminus S_{T,\alpha}$. In particular, F_α is piecewise affine outside of $S_{B,\alpha} \cup S_{T,\alpha}$. $\square$

Lemma 7 gives rise to a characterization of the modes of the derivative of a taut string (see Lemma 10 below). This characterization resembles the fact that an isolated local maximum (local minimum) of f_α corresponds to a point (or interval) where its antiderivative F_α changes from being locally convex to locally concave (concave to convex), see Fig. 5. Accordingly, we define:

Definition 3 (maximally concave, convex, and affine intervals) Fix $\alpha > 0$. An interval $I = [a, b] \subset [0, 1]$ is called *maximally affine* if F_α is affine on I but not on any interval that properly contains I . An interval $I = [a, b] \subset [0, 1]$ that is *not* maximally affine is called *maximally convex* (concave) if F_α is convex (concave) on I but not on any interval that properly contains I . $\square$

Observe that by Lemma 7, if F_α is not affine on all of $[0, 1]$, then every $t \in [0, 1]$ is contained in a maximally concave or a maximally convex interval (or possibly both). By construction, maximally convex (concave) intervals are mutually disjoint (within their respective classes).

Definition 4 (positive and negative inflection intervals) Fix $\alpha > 0$. An interval $I = [a, b] \subset (0, 1)$ is called a *positive* (*negative*) *inflection interval* of F_α if I is a maximally affine interval of F_α and F_α is convex (concave) on some non empty neighborhood of a and concave (convex) on some non empty neighborhood of b . $\square$

Notice that we deliberately require that $a > 0$ and $b < 1$ in our definition of inflection intervals. As a direct consequence of Lemma 7 we obtain:

Lemma 8 Fix $\alpha > 0$. If $[a, b]$ is a positive inflection interval of F_α , then $F_\alpha(a) = T_\alpha(a)$ and $F_\alpha(b) = B_\alpha(b)$; if it is a negative inflection interval, then $F_\alpha(a) = B_\alpha(a)$ and $F_\alpha(b) = T_\alpha(a)$. $\square$

Moreover we have:

Lemma 9 Fix $\alpha > 0$. Then F_α has the following properties:

- (i) The number of maximally convex, the number of maximally concave, and the number of inflection intervals of F_α is finite.
- (ii) Maximally convex and maximally concave intervals are interleaved, i.e., the set of points between two consecutive maximally convex (concave) intervals belongs to a maximally concave (convex) interval.
- (iii) The intersection of a maximally convex (concave) with an immediately consecutive maximally concave (convex) interval is a positive (negative) inflection interval, and every inflection interval arises in this way. $\square$

PROOF Let T_α and B_α denote the top and bottom of the α -tube around F_α , respectively. Since F_α is continuous, the graphs of T_α and B_α are compact sets. Let I be a maximally concave, a maximally convex, or an inflection interval of F_α . By Definitions 3 and 4 and Lemma 7, the graph of F_α restricted to I must then contain an affine segment that connects T_α with B_α (or B_α with T_α). Therefore, the arc length of the graph of F_α , restricted to I , is bounded from below by the

Euclidean distance d_α between the graphs of B_α and T_α . Since these sets are compact and disjoint, one has $d_\alpha > 0$. Since d_α is independent of I , and since F_α is Lipschitz, it follows that the length of I is bounded from below by a number that only depends on α and the Lipschitz constant of F_α . Hence, since maximally convex (concave) intervals are mutually disjoint, there can only exist finitely many of them. Likewise, since positive (negative) inflection intervals are disjoint, there can only exist finitely many of those. Properties (i) and (ii) are then a straightforward consequence of Lemma 7. ■

The next lemma states the promised characterization of the modes of the derivative of a taut string.

Lemma 10 *Fix $\alpha > 0$ and define*

$$f_\alpha(t) = \lim_{\epsilon \rightarrow 0} \inf_{0 < \delta < \epsilon} \frac{F_\alpha(t + \delta) - F_\alpha(t - \delta)}{2\delta} \quad \text{if } 0 < t < 1$$

and $f_\alpha(t) = \lim_{s \rightarrow t} f_\alpha(s)$ for $t \in \{0, 1\}$. Then the number of positive inflection intervals of F_α equals the number of modes of f_α , and this number is finite. □

PROOF First notice that the definition of $f_\alpha(0)$ and $f_\alpha(1)$ is meaningful since F_α is affine in some neighborhood of 0 and 1.

If F_α is affine on all of $[0, 1]$, then there is nothing to show. So suppose that this is not the case. Consider a finite partition $P = \{t_0, \dots, t_{|P|}\}$ of $[0, 1]$. Notice that f_α is nowhere decreasing (nowhere increasing) on intervals where F_α is convex (concave). Hence, for t_i to count a mode of f_α , i.e., $\mathcal{M}(f_\alpha, P, i) = 1$, the pair (t_{i-1}, t_i) must not belong to the same maximally concave interval and the pair (t_i, t_{i+1}) must not belong to the same maximally convex interval of F_α . Since, by assumption, F_α is not affine on all of $[0, 1]$, every $t \in [0, 1]$ belongs to a maximally concave or maximally convex interval (or both). Therefore, by property (i) of Lemma 9, to each mode of f_α counted by P there corresponds at least one change from a maximally convex to an immediately consecutive maximally concave interval. By property (ii) of Lemma 9, the total number of such changes is equal to the number of positive inflection intervals, which we denote by $\mathcal{J}^+(F_\alpha)$. It follows that $\mathcal{J}^+[F_\alpha] \geq \mathcal{M}(f_\alpha)$.

Vice versa, by considering a partition of $[0, 1]$ such that there exists (apart from $t_0 = 0$ and $t_{|P|} = 1$) exactly one point in each positive and each negative inflection interval, it is straightforward to show that $\mathcal{J}^+[F_\alpha] \leq \mathcal{M}(f_\alpha)$.

Finally, finiteness of $\mathcal{M}(f_\alpha)$ follows from the fact that there are only finitely many positive inflection intervals. ■

With these preparations, we are now in the position to prove Theorem 8.

PROOF (OF THEOREM 8) Let $g \in \mathcal{L}$ with antiderivative G such that $d_K(f, g) \leq \alpha$. Consider a positive inflection interval $[a, b]$ of F_α . By Lemma 8, $F_\alpha(a) = T_\alpha(a)$, $F_\alpha(b) = B_\alpha(b)$, and F_α is affine on $[a, b]$. In particular, $G(a) \leq F_\alpha(a)$ and $G(b) \geq F_\alpha(b)$, and thus

$$f_\alpha(t) = \frac{F_\alpha(b) - F_\alpha(a)}{b - a} \leq \frac{G(b) - G(a)}{b - a} \quad \text{for all } t \in (a, b).$$

For every Lebesgue-integrable $g : [a, b] \rightarrow \mathbb{R}$ with $G(t) = G(a) + \int_a^t g(s) ds$, there exist sets $C_1, C_2 \subset [a, b]$ of positive Lebesgue measure such that

$$g(c_1) \leq \frac{G(b) - G(a)}{b - a} \leq g(c_2)$$

for all $c_1 \in C_1$ and all $c_2 \in C_2$.

Hence, for every positive inflection interval $[a, b]$ there exists $t \in (a, b)$ such that $g(t) \geq f_\alpha(t)$. By a similar argument, for every negative inflection interval $[a, b]$ there exists $t \in (a, b)$ such that $g(t) \leq f_\alpha(t)$. By Lemma 10, whenever $\mathcal{M}(f_\alpha) > 0$ (otherwise there is nothing to show), the set of positive inflection intervals of F_α is not empty. Therefore, one can choose a partition $P = \{t_0, \dots, t_{|P|}\}$ of $[0, 1]$ that contains (apart from $t_0 = 0$ and $t_{|P|} = 1$) exactly one point in the interior of each inflection interval of F_α such that $g(t_i) \geq f_\alpha(t_i)$ whenever t_i lies in a positive inflection interval and $g(t_i) \leq f_\alpha(t_i)$ whenever t_i lies in a negative inflection interval. By the proof of Lemma 10, $\mathcal{M}(f_\alpha) = \mathcal{M}(f_\alpha, P)$ for any partition P that contains (apart from $t_0 = 0$ and $t_{|P|} = 1$) exactly one point in the interior of each inflection interval. Such partitions P count a mode of f_α precisely for every *positive* inflection interval of F_α . Since positive and negative inflection intervals are interleaved and their interiors are disjoint, we obtain that $\mathcal{M}(g, P) \geq \mathcal{M}(f_\alpha, P) = \mathcal{M}(f_\alpha)$. Thus $\mathcal{M}(g) \geq \mathcal{M}(f_\alpha)$. ■

6 Computing Kolmogorov signatures

The results of the previous section lead to an efficient algorithm for computing Kolmogorov signatures. Let $X \subset \mathcal{L}$ be some subset, and let $f \in X$ with antiderivative F . Suppose that X contains the derivatives f_α of the taut strings F_α for all $\alpha \geq 0$. For example, let X be the space of piecewise constant functions. For α large enough, F_α is affine on all of $[0, 1]$, and its derivative f_α has no modes. If f has any modes at all, then by lowering α continuously, F_α will at some point develop a positive inflection interval below some threshold $\alpha_0 > 0$. By Theorem 8 and Lemma 10, the value of α_0 is precisely the distance of f to the set of functions in X with zero modes, i.e., $s_0(f) = \alpha_0$. Continuing this way, and defining α_k as the smallest α for which f_α has at most k modes, one finds that $s_k(f) = \alpha_k$ for all k .

The idea of the algorithm below is to reverse this observation: Starting from $f = f_0$, we incrementally compute the values of α (in *increasing* order) at which the number of modes of f_α decreases. To this end, we work with the space X of piecewise constant functions on a fixed partition $0 = t_0 < t_1 < \dots < t_n = 1$ of $[0, 1]$. Notice that since we require $X \subset \mathcal{L}$, we have $f(t_i) = \frac{1}{2}(f|_{(t_{i-1}, t_i)} + f|_{(t_i, t_{i+1})})$ for all non-boundary points t_i of the partition.

Our starting point is a reformulation of Lemma 7 for piecewise constant functions.

Lemma 11 *Let f be a piecewise constant function with antiderivative F . Then the taut string F_α is the unique continuous piecewise linear function in $D_\alpha(F)$ with $F_\alpha(0) = F(0)$ and $F_\alpha(1) = F(1)$ such that if t is an increasing (decreasing) discontinuity of $f_\alpha = F'_\alpha$, then $F_\alpha(t) = F(t) + \alpha$ ($F_\alpha(t) = F(t) - \alpha$).* □

Fix $\alpha \geq 0$. Let $I = (a, b) \subseteq (0, 1)$ be an open interval, and let f_α be constant on (a, b) . We call I *regular* for f_α if either $a = 0$ and $b = 1$ or $a > 0$, $b < 1$, and there exists $\epsilon > 0$ such that for all $0 < \delta \leq \epsilon$ either $f_\alpha(a) > f_\alpha(a - \delta)$ and $f_\alpha(b) < f_\alpha(b + \delta)$ or $f_\alpha(a) < f_\alpha(a - \delta)$ and $f_\alpha(b) > f_\alpha(b + \delta)$. We call $I = (a, b)$ *maximal (respectively minimal)* for f_α if $a > 0$, $b < 1$, and there exists $\epsilon > 0$ such that for all $0 < \delta \leq \epsilon$ one has $f_\alpha(a) > f_\alpha(a - \delta)$ and $f_\alpha(b) > f_\alpha(b + \delta)$ (respectively $f_\alpha(a) < f_\alpha(a - \delta)$ and $f_\alpha(b) < f_\alpha(b + \delta)$). We call I *critical* if it is minimal or maximal. Finally, we call $I = (a, b)$ a *boundary interval* for f_α if either $a = 0$ and $b < 1$ or $a > 0$ and $b = 1$, and (a, b) is the largest such interval on which f_α is constant. As a consequence of Lemma 11 we obtain:

Corollary 2 *Away from discontinuities, f_α has the following form: either*

- t lies on a regular interval $I = (a, b)$ of f_α with value $f_\alpha(t) = \frac{F(b)-F(a)}{b-a}$,
- t lies on a locally minimal/maximal interval $I = (a, b)$ of f_α with value $f_\alpha(t) = \frac{F(b)-F(a) \pm 2\alpha}{b-a}$,
or
- t lies on a boundary interval of f_α with value $f_\alpha(t) = \frac{F(b)-F(a) \pm \alpha}{b-a}$. $\square$

This corollary is central for our computation of Kolmogorov signatures. First, observe that a value of a maximal interval is continuously decreasing with growing α , the value of a minimal interval is continuously increasing, and the value of a regular interval remains unchanged. Moreover, if α is increased only slightly, then the discontinuities of f_α remain unchanged; indeed:

Lemma 12 *Let F be piecewise linear. For every $\alpha \geq 0$ there is $\delta > 0$ such that the points of discontinuity of f_β coincide with those of f_α for all β with $\alpha \leq \beta < \alpha + \delta$. Moreover, if t lies on a regular interval of f_α , then $f_\beta(t) = f_\alpha(t)$. $\square$*

PROOF Define G_β by the properties of Lemma 11, using the discontinuities of f_α , i.e., if t is an increasing (decreasing) discontinuity of f_α , then define $G_\beta(t) := F(t) + \beta$ (resp. $G_\beta(t) := F(t) - \beta$); set $G_\beta(0) := F(0)$ and $G_\beta(1) := F(1)$, and interpolate linearly. Then $\|G_\beta - F_\alpha\|_\infty = \beta - \alpha$ and thus, since $F_\alpha \in D_\alpha(F)$, we have that $\|G_\beta - F\|_\infty \leq \beta$, i.e., $G_\beta \in D_\beta(F)$. For δ sufficiently small, the discontinuities of $g_\beta = G'_\beta$ have the same type as those of f_α . But since F_β is uniquely defined by the properties of Lemma 11 with respect to these discontinuities, we must have $F_\beta = G_\beta$. $\blacksquare$

As a consequence, for every $\alpha \geq 0$, there exists a minimal number $\mu(\alpha) > \alpha$ such that f_β and f_α have the same points of discontinuity for all β with $\alpha \leq \beta < \mu(\alpha)$ but the set of points of discontinuity of $f_{\mu(\alpha)}$ is different from that of f_α . We call $\mu(\alpha)$ the *merge value* of α . The merge value is the smallest number strictly greater than α for which a critical interval or a boundary interval of $f_{\mu(\alpha)}$ reaches the value of an adjacent constant interval, and the corresponding discontinuity vanishes. Each discontinuity of f_α that is incident to a critical or a boundary interval is a possible candidate for such an event. Consider such a discontinuity b between two consecutive constant intervals $I = [a, b]$ and $J = [b, c]$ of f_α . For an interval $I = [a, b]$, let $F_I := F(b) - F(a)$. As a consequence of Corollary 2 and Lemma 12, we obtain that the merge value $\mu(\alpha)$ is the smallest number among all merge value *candidates* $m_{I,J}$ of f_α , which are computed as follows:

If I is critical and J is regular or vice-versa, then the merge value candidate is

$$m_{I,J} = \frac{1}{2} \left| F_I - \frac{|I|}{|J|} F_J \right|.$$

If both I and J are critical, then the merge value candidate is

$$m_{I,J} = \left| \frac{|I|F_J - |J|F_I}{2(|I| + |J|)} \right|.$$

If I is critical and J is a boundary interval, then the merge value candidate is

$$m_{I,J} = \left| \frac{|I|F_J - |J|F_I}{|I| + 2|J|} \right|.$$

If I is a boundary interval and J is critical, then the merge value candidate is

$$m_{I,J} = \left| \frac{|I|F_J - |J|F_I}{2|I| + |J|} \right|.$$

If I is a boundary interval and J is regular or vice-versa, then the merge value candidate is

$$m_{I,J} = \left| F_I - \frac{|I|}{|J|} F_J \right|.$$

If both I and J are boundary intervals, then the merge value candidate is

$$m_{I,J} = \left| \frac{|I|F_J - |J|F_I}{|I| + |J|} \right|.$$

We define the sequence of merge values $\mu_1 < \mu_2 < \mu_3 < \dots$ of f as follows. Starting from $\mu_1 := \mu(0)$, let $\mu_{i+1} := \mu(\mu_i)$. By construction, the values $\alpha = \mu_i$ are precisely those values where the number of discontinuities of f_α decreases with increasing α .

Observe that the merge value *candidates* of $f_{\mu_{i+1}}$ are equal to those of f_{μ_i} *except only* for the merged intervals I and J , i.e., those intervals that have the same value for $f_{\mu_{i+1}}$ but did not have the same value for f_{μ_i} . This suggests an efficient way for computing Kolmogorov signatures of f in reverse order. Starting with $\alpha = 0$, we iterate in increasing order through the sequence of merge values of f . In a min-priority queue, we maintain the merge value *candidates* $m_{I,J}$. In each iteration i , the lowest merge value candidate is the next value μ_i . Upon a merge, the corresponding discontinuity is removed, and the merge value candidates of the neighboring discontinuities are recomputed and updated in the priority queue. The discontinuities are organized in a linked list to allow fast access to the neighbors. If the number of modes of f_α has decreased upon a merge, the value α is prepended to the sequence of computed signatures. This can only occur if one of the merged intervals is maximal. The method is summarized in pseudocode in Algorithm 1. Using an appropriate heap data structure, the running time is $O(n \log n)$, where n is the number of function values of f .

Algorithm 1 Computing Kolmogorov signatures

```

1: procedure KOLMOGOROVSEQUENCE( $f$ : list of function values)
2:    $\alpha = 0$ 
3:    $S =$  empty sequence
4:    $L =$  jumps of  $f$  (linked list)
5:    $Q =$  merge values of the jumps (priority queue)
6:   while the priority queue  $Q$  is not empty do
7:     let  $\alpha$  be the smallest merge value in  $Q$ 
8:     let  $I = [a, b]$  and  $J = [b, c]$  be the corresponding intervals
9:     if  $I$  and  $J$  are minimum/maximum or boundary/maximum of  $f_\alpha$  then
10:       prepend  $\alpha$  to  $S$ 
11:       remove  $b$  from the list  $L$  of discontinuities
12:       remove  $\alpha$  from the priority queue  $Q$ 
13:       recompute merge values of  $a$  and  $c$  and update priority queue  $Q$ 
14:   return  $S$ 

```

Acknowledgements

We would like to thank the anonymous reviewers for their very helpful suggestions for revising our manuscript and Carola Schoenlieb for inspiring discussions.

References

- [1] Sivaraman Balakrishnan, Alessandro Rinaldo, Don Sheehy, Aarti Singh, and Larry A. Wasserman, *Minimax rates for homology inference*, Journal of Machine Learning Research - Proceedings Track **22** (2012), 64–72.
- [2] Ulrich Bauer, Carsten Lange, and Max Wardetzky, *Optimal topological simplification of discrete functions on surfaces*, Discrete & Computational Geometry **47** (2012), no. 2, 347–377.
- [3] Paul Bendich, Taras Galkovskyi, and John Harer, *Improving homology estimates with random walks*, Inverse Problems **27** (2011), no. 12, 124002+.
- [4] Patrick Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication.
- [5] Peter Bubenik, Gunnar Carlsson, Peter T. Kim, and Zhi-Ming Luo, *Statistical topology via Morse theory persistence and nonparametric estimation*, Algebraic Methods in Statistics and Probability II (Marlos A. G. Viana and Henry P. Wynn, eds.), Contemporary Mathematics, vol. 516, American Mathematical Society, 2010, pp. 75–92.
- [6] Peter Bubenik and Peter T. Kim, *A statistical approach to persistent homology*, Homology, Homotopy and Applications **9** (2007), no. 2, 337–362.
- [7] Hock Peng Chan and Guenther Walther, *Detection with the scan and the average likelihood ratio*, Statistica Sinica **23** (2013), 409–423.
- [8] Frédéric Chazal, David Cohen-Steiner, Leonidas J. Guibas, Facundo Mémoli, and Steve Y. Oudot, *Gromov-Hausdorff stable signatures for shapes using persistence*, Computer Graphics Forum **28** (2009), no. 5, 1393–1403.
- [9] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot, *Geometric inference for probability measures*, Foundations of Computational Mathematics **11** (2011), no. 6, 733–751.
- [10] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, *Stability of persistence diagrams*, Discrete and Computational Geometry **37** (2007), no. 1, 103–120.
- [11] P. L. Davies and A. Kovac, *Local extremes, runs, strings and multiresolution*, The Annals of Statistics **29** (2001), no. 1, 1–65.
- [12] P. Laurie Davies and Arne Kovac, *Densities, spectral densities and modality*, The Annals of Statistics **32** (2004), no. 3, 1093–1136.
- [13] David Donoho and Jiashun Jin, *Higher criticism for detecting sparse heterogeneous mixtures*, Ann. Statist. **32** (2004), no. 3, 962–994.
- [14] David L. Donoho, *One-sided inference about functionals of a density*, The Annals of Statistics **16** (1988), no. 4, 1390–1420.

- [15] David L. Donoho, Iain M. Johnstone, Gérard Kerkycharian, and Dominique Picard, *Wavelet shrinkage: asymptopia?*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 2, 301–369.
- [16] Herbert Edelsbrunner and John L. Harer, *Computational topology: An introduction*, AMS, 2010.
- [17] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian, *Topological persistence and simplification*, Discrete and Computational Geometry **28** (2002), no. 4, 511–533.
- [18] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh, *Confidence sets for persistence diagrams*, Ann. Statist. **42** (2014), no. 6, 2301–2339.
- [19] I. J. Good and R. A. Gaskins, *Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data*, Journal of the American Statistical Association **75** (1980), no. 369, pp. 42–56.
- [20] Markus Grasmair, *The equivalence of the taut string algorithm and BV-regularization*, Journal of Mathematical Imaging and Vision **27** (2007), no. 1, 59–66.
- [21] Markus Grasmair and Andreas Obereder, *Generalizations of the taut string method*, Numerical Functional Analysis and Optimization **29** (2008), no. 3-4, 346–361.
- [22] J. A. Hartigan and P. M. Hartigan, *The dip test of unimodality*, The Annals of Statistics **13** (1985), no. 1, pp. 70–84.
- [23] John A. Hartigan, *Testing for antimodes*, Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin Heidelberg, 2000, pp. 169–181.
- [24] Yuri Ingster and I.A. Suslina, *Nonparametric goodness-of-fit testing under gaussian models*, Lecture Notes in Statistics, vol. 169, Springer, 2003.
- [25] Jennifer Kloeke and Gunnar Carlsson, *Topological De-Noising: Strengthening the topological signal*, February 2010, arXiv:0910.5947.
- [26] Enno Mammen and Sara van de Geer, *Locally adaptive regression splines*, The Annals of Statistics **25** (1997), no. 1, 387–413.
- [27] Emmanuel Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 31, Springer-Verlag, Berlin, 2000.
- [28] Donald R. Sheehy, *A multicover nerve for geometric inference*, CCCG: Canadian Conference in Computational Geometry, 2012, pp. 309–314.
- [29] Galen R. Shorack and Jon A. Wellner, *Empirical processes with applications to statistics*, Classics in Applied Mathematics, vol. 59, Siam, 2009.
- [30] B. W. Silverman, *Using kernel density estimates to investigate multimodality*, Journal of the Royal Statistical Society. Series B (Methodological) **43** (1981), no. 1, pp. 97–99.
- [31] A. W. van der Vaart, *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2000.

[32] M.P. Wand and M.C. Jones, *Kernel smoothing*, London: Chapman & Hall, 1995.