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

Ulrich Bauer\textsuperscript{1}, Axel Munk\textsuperscript{2,3}, Hannes Sieling\textsuperscript{2}, and Max Wardetzky\textsuperscript{4}

\textsuperscript{1}Technische Universität München (TUM)
\textsuperscript{2}Institute for Mathematical Stochastics, University of Göttingen
\textsuperscript{3}Max Planck Institute for Biophysical Chemistry, Göttingen
\textsuperscript{4}Institute of Numerical and Applied Mathematics, University of Göttingen

January 27, 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; secondary 62H12

1 Introduction

Persistent homology \cite{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 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 \cite{17}. In this article, we consider what we call persistence signatures, defined as (half) the life span (or persistence) of critical values, i.e.,

\textsuperscript{*}Research partially supported by DFG FOR 916, Volkswagen Foundation, and the Torosys project FP7-ICT-318493-STREP
persistence signatures correspond to (half) the lengths of persistence barcodes and (when properly ordered) give rise to a descending sequence

\[ s_{0,\infty}(f) \geq s_{1,\infty}(f) \geq s_{2,\infty}(f) \geq \cdots , \tag{1} \]

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 \( f : [0, 1] \to \mathbb{R} \). 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 \( f : [0, 1] \to \mathbb{R} \) with value less or equal to \( \delta \), it suffices to change \( f \) by \( \delta \) 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 \textit{Kolmogorov metric} \( d_K \) and its resulting Kolmogorov signatures \( s_K \). For one dimensional signals \( f : [0, 1] \to \mathbb{R} \) the Kolmogorov norm is defined as the \( L^\infty \)-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 \textit{mode hunting}, a widely studied problem in statistics \([23, 19, 22, 12, 30]\), and the robust estimation of signatures. Most related to our approach is \([12]\), where the Kolmogorov norm has been used for mode hunting in the context of density estimation.

\(^1\)Note that this result does no longer hold in dimensions greater than two.
In the sequel we consider the following basic statistical additive regression model. Suppose that \( f : [0, 1] \rightarrow \mathbb{R} \) is corrupted by random noise \( \epsilon \) and observed by a finite number of (equidistantly sampled) measurements \((Y_i)_{i=0}^n\), i.e.,

\[
Y_i = f(t_i) + \epsilon_i, \quad t_i = i/n.
\]

(2)

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

\[
\mathbb{E}|\epsilon_i|^m \leq \nu m!\kappa^{m-2}/2 \quad \text{for all } i = 1, \ldots, n.
\]

(3)

We are concerned with the following question: \textit{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 \([1, 3, 5, 6, 8, 9, 25, 28]\). 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., \([1, 6, 18]\) for the case of a kernel estimator. In fact, presmoothing with a kernel estimator leads to what has been sometimes called the \textit{notorious bandwidth selection problem}, which does not posses 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

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

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

$$\mathbb{P}\left( s_k(f) \in \left[ (s_k(Y) - \tau_n(\alpha))^+, s_k(Y) + \tau_n(\alpha) \right] \right. \text{ for all } k \in \mathbb{N}_0 \geq 1 - \alpha,$$

where $(x)^+ = \max(0, x)$. Here $\tau_n(\alpha)$ depends in an explicit manner on $n$, $\alpha$, $\kappa$, and $v$, which are known constants or can be easily estimated from the data. We drop the dependence of $\tau_n$ on $\kappa$ and $v$ by considering $\kappa$ and $v$ fixed since we are mainly concerned with the dependence on $n$ and $\alpha$. For fixed $\alpha, \kappa, v$, 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

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

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
\[
P(k_{\tau_n(\alpha)}(f) < k_{2\tau_n(\alpha)}(f)) \leq \alpha.
\]
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
\[
P(k_{2\tau_n(\alpha/2)}(f) \leq k_{\tau_n(\alpha/2)}(Y) \leq k) \geq 1 - \alpha.
\]
As mentioned before, for fixed \( \alpha, \kappa, v \), 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 \( \alpha \) 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
\[
P(k_{\epsilon/2}(Y) = k) \geq 1 - 2 \exp\left(-\frac{\epsilon^2 n}{8v + 4k\epsilon}\right), \tag{5}
\]
i.e., the number of modes of \( f \) can be estimated exponentially fast (in the number of samples) 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 \( \epsilon_i \). 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.

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
\[
f_n(x) = \begin{cases} (1 + \epsilon) \sqrt{2 \log n} & \text{if } x \in [j/n, (j + 1)/n), \\ 0 & \text{otherwise}, \end{cases} \tag{6}
\]
for some \( \epsilon > 0 \) and for some \( j \in \{0, \ldots, n - 1\} \) that is a priori not known. We show in Theorem 6 that asymptotically (as \( n \to \infty \)) 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 \((Y_1, \ldots, Y_n)\) is known to behave asymptotically minimax efficient in the sense of detecting a non vanishing mode with probability tending to one as \( n \to \infty \), 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
\[
f_n(x) = \begin{cases} \delta_n & \text{if } x \in [1/3, 2/3), \\ 0 & \text{otherwise}, \end{cases} \tag{7}
\]
with $\delta_n \to 0$. It is well known that it is possible to detect the single mode of $f_n$ with probability tending to one as $n \to \infty$ if $\delta_n \sqrt{n} \to \infty$, 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 $\delta_n \sqrt{n} \to \infty$, and assume additionally $\delta_n \sqrt{\log n} \to 0$. Then Theorem 7 asserts that for an arbitrary sequence $(q_n)$ of reals one has $\limsup_{n \to \infty} P(k_{q_n}(\bar{Y}) = 1) < 1$.

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, $F_\alpha$, is the function whose graph has minimal total length (as a curve) among all absolutely continuous functions in the $\alpha$-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: $f_\alpha$ minimizes the number of modes among all $L^\infty$-functions in the (closed) $\alpha$-ball around $f$ with respect to the Kolmogorov norm. Indeed, our result generalizes previous results on the mode-minimizing property of $f_\alpha$, which were shown in the special context of piecewise constant functions using the Kolmogorov norm, see, e.g., [11, 12, 22, 26].

2 Modes and signatures

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

$$X_k := \{ f \in X : M(f) \leq k \}$$

Define the $k$th metric signature of $f \in X$ as

$$s_k(f) := d(f, X_k) = \inf_{g \in X_k} d(f, g)$$

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 $M(f) > k$ but $s_k(f) = d(f, X_k) = 0$. Hence:

Definition 2 (Descriptive metric) $(X, d)$ is called descriptive if $M(f) > k$ 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 $(X, d_\infty)$ is descriptive for every $X \subset \mathcal{L}^{\infty}$. □
which proves the first claim. 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., $H_x(f^{-1}(-\infty, t))$ 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 $d_{\infty}$ 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 $|p_k(g) - p_k(f)| \leq 2d_{\infty}(f, g)$ 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 $\delta$ in the sup norm, it is possible to remove all its persistence pairs with persistence less or equal to $2\delta$ 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

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

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

\[8\]
We thus obtain a projection operator $P$ for all $0 \leq \delta$. Let $\delta > 0$ and assume that $t \in \mathbb{R}$ exists $t \in [0, 1]$ and $f \in L^\infty$, and since Lebesgue’s differentiation theorem asserts that every $f \in L^1$ satisfies

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

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

The following lemma further motivates our choice of $L$.

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

**Proof** Let $f \in L^\infty$ with representative $\tilde{f} := P(f) \in L$. 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., $|\tilde{f}(t_i) - \max(f(t_{i-1}), f(t_{i+1}))| > \epsilon$ for some $\epsilon > 0$. Consider any open neighborhood $U_i$ of $t_i$. Since

$$\tilde{f}(t_i) = \lim_{\epsilon \to 0} \inf_{0 < \delta < \epsilon} \frac{1}{2\delta} \int_{t_i-\delta}^{t_i+\delta} f(s) \, ds,$$

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 $f(t_i') \leq f(t_{i-1}) + \epsilon/2$ and $f(t_i') \leq f(t_{i+1}) + \epsilon/2$. By our choice of $\epsilon$ this implies $f(t_i') > \max(f(t_{i-1}), f(t_{i+1}))$. Continuing this way yields a partition $P'$ with $\mathcal{M}(f, P') \geq \mathcal{M}(\tilde{f}, P)$.

**Lemma 5** $(X, d_k)$ is descriptive for every $X \subset L$.

**Proof** We show that $X \setminus X_k$ is open wrt. the Kolomogorov metric. Let $f \in X \setminus X_k$. Then there exists a finite partition $P = \{t_0, \ldots, t_{|P|}\}$ of $[0, 1]$ and some index set $I$ with $k < |I| < |P|$ such 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 $d_k(f, g) < \frac{\epsilon}{4} \delta'$. Then

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

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

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

9
for all $i \in I$ and all $j \in \{-1,0,1\}$. Therefore, there exists $t'_i \in [t_i - \delta_i, t_i + \delta_i]$ with $g(t'_i) > f(t_i) - \frac{\epsilon}{2}$. Likewise, there exist $t'_{i+1} \in [t_{i+1} - \delta_{i+1}, t_{i+1} + \delta_{i+1}]$ with $g(t'_{i+1}) < f(t_{i+1}) + \frac{\epsilon}{2}$. Thus there exists a partition $P'$ of $[0,1]$ with $\mathcal{M}[P'] > k$, i.e., $g \in X \setminus X_k$. Since $\epsilon$ and $\delta'$ 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 $(\epsilon_i)$ in Model (2) is independently distributed with mean zero such that for some $\kappa > 0$, $\nu > 0$ and all $m \geq 2$,

$$
\mathbb{E}[|\epsilon|^m] \leq \nu m! \kappa^{m-2}/2 \quad \text{for all } i = 1, \ldots, n .
$$

Distributions which satisfy (9) include the centered normal distribution with variance $\sigma^2 > 0$, the (centered) Poisson distribution with intensity $\lambda$, or the Laplace distribution with variance $2\lambda^2$. Moreover, any symmetric distribution around zero with compact support is covered by (9), including the uniform distribution on an interval $[-B,B]$.

### 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 $(I^m(\alpha))_{j \in \mathbb{N}_0}$ 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 $f^{(n)} : [0, 1] \to \mathbb{R}$ defined as

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

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

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

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

$$d_K(f, f^{(n)}) = \sup_{x \in [0,1]} \left| \int_0^x f(t) - f^{(n)}(t) dt \right|. \quad (10)$$

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.,

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

then

$$d_K(f, f^{(n)}) \leq \frac{C}{\gamma+1} n^{-\gamma}, \quad (11)$$

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 (9). Then, for any $\delta > 0$ and any $f \in X$, one has
\[
\mathbb{P}\left(\max_{j \geq 0} |s_j(Y^{(n)}) - s_j(f^{(n)})| \geq \delta \right) \leq 2 \exp\left(-\frac{\delta^2 n}{2v + 2\kappa \delta}\right). \tag{\text{\{\textit{proof}\}}}
\]

**Proof** By stability of metric signatures (Lemma 1), we have that
\[
\mathbb{P}\left(\max_{j \geq 0} |s_j(Y^{(n)}) - s_j(f^{(n)})| \geq \delta \right) \leq \mathbb{P}\left(d_K(f^{(n)}, Y^{(n)}) \geq \delta \right).
\]

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
\[
\mathbb{P}\left(d_K(Y^{(n)}, f^{(n)}) \geq \delta \right) = \mathbb{P}\left(\max_k |S_k| \geq \delta n \right) \leq 2 \exp\left(-\frac{\delta^2 n}{2v + 2\kappa \delta}\right). \tag{\text{\{\textit{proof}\}}}
\]

This result 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.

A useful application of Theorem 1 is that for a given probability $\alpha$, 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.

**Theorem 2** Fix some $\alpha \in (0, 1)$ and let
\[
\tau_n(\alpha) := \frac{1}{n} \left(\sqrt{\log(\alpha/2)(\log(\alpha/2)\kappa^2 - 2\nu\kappa)} - \kappa \log(\alpha/2)\right).
\]

Assume the regression model (2) and the moment condition in (9). Then
\[
\inf_{f \in X} \mathbb{P}\left(s_j(f^{(n)}) \in \left[(s_j(Y^{(n)}) - \tau_n(\alpha))^+, s_j(Y^{(n)}) + \tau_n(\alpha)\right] \quad \text{for all } j \in \mathbb{N}_0 \right) \geq 1 - \alpha,
\]
where $(x)_+ = \max(0, x)$.

**Proof** From Theorem 1 we obtain
\[
\mathbb{P}\left(|s_j(Y^{(n)}) - s_j(f^{(n)})| \leq \tau_n(\alpha) \quad \text{for all } j \in \mathbb{N}_0 \right) \geq 1 - 2 \exp\left(-\frac{\tau_n(\alpha)^2 n}{2\nu + 2\kappa \tau_n(\alpha)}\right)
\]
\[
= 1 - \alpha.
\]

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

Note that $\tau_n(\alpha)$ is a quantity that only depends on the values $n, \kappa, \nu$, and the confidence level $\alpha$. Here we assume for simplicity that $\kappa$ and $\nu$ 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 $\sigma^2$. Fixing $\alpha$, we obtain a (random) sequence of intervals

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

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 $\alpha$, 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 $n \to \infty$.

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

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

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

13
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 (9), let $\alpha \in (0, 1)$, and let $k \in \mathbb{N}_0$ be such that $f^{(n)} \in X_k$. Then

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

**Proof** First, observe from the definition of $k_{\tau_n(\alpha)}(Y^{(n)})$ in (12) 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

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

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 $\alpha$. 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 (12), we find:

**Theorem 4** Let $f \in X$, assume the regression model (2) and the moment condition (9), let $\alpha \in (0, 1)$, and let $k \in \mathbb{N}_0$ be such that $f^{(n)} \in X_k$. Then

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

**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 (13) is trivially satisfied, since $k_{\tau_n(\alpha)}(Y^{(n)}) \geq 0$. So suppose that $l > 0$. Then

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

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

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 $\alpha, \kappa, \nu$ 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 $\alpha$ that all signatures above this threshold are detected.

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

**Corollary 1** Assume the regression model (2), the moment condition (9), let $\varepsilon \geq 0$, and let $f^{(n)} \in X_k$. Define

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

and

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

Then

$$\mathbb{P}(k_\varepsilon(f^{(n)}) \in [l(\alpha, \varepsilon), u(\alpha, \varepsilon)]) \geq 1 - \alpha.$$  

**Proof** Suppose, for the moment, that

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

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

Further, while still assuming (14), $s_j(Y^{(n)}) > \varepsilon + \tau_n(\alpha)$ implies that $s_j(f^{(n)}) > \varepsilon$. Hence, by the definition of $l(\alpha, \varepsilon)$, we find that $s_l(\alpha, \varepsilon)(f^{(n)}) > \varepsilon$. This in turn implies $k_\varepsilon(f^{(n)}) \geq l(\alpha, \varepsilon)$. Therefore, we have so far shown that (14) implies that

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

Since (14) holds with probability $\geq 1 - \alpha$ (see proof of Theorem 1), this proves the assertion. ■

Note that the upper bound for $k_\varepsilon$ jumps to $\infty$ if $\varepsilon \leq \tau_n(\alpha)$. 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

$$\sqrt{n} \left( s_k(Y^{(n)}) - s_k(f^{(n)}) \right) = \sqrt{n} \inf_{g \in X_k} \sup_{x \in [0,1]} \left| \int_0^x \left( \epsilon^{(n)}(t) - g(t) \right) dt \right| = \inf_{g \in X_k} \sup_{x \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{[nx]} \epsilon_i - \int_0^x \sqrt{n} g(t) dt \right|$$

$$= \inf_{g \in X_k} \sup_{x \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{[nx]} \epsilon_i - \int_0^x g(t) dt \right|.$$
where the last equality follows from the scaling invariance of $X_k$. Noting that
\[
\frac{1}{\sqrt{n}} \max_{m=1,\ldots,n} \left| \sum_{i=1}^{m} e_i \right| \overset{D}{\to} \sup_{0 \leq s \leq 1} B(x),
\]
where $B$ denotes a standard Brownian motion on $[0,1]$ and using that $f \equiv 0$, it follows that
\[
\sqrt{n} \left( s(Y^{(n)}) \right) \overset{D}{\to} s(B'),
\]
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.  

\begin{remark}[Gaussian observation] If the noise $\epsilon$ in (2) is Gaussian with mean zero and variance $\sigma^2$, then Theorem 1 can be sharpened, due to a refined large deviation result for Gaussian observations (see, e.g., [4]):
\[
\mathbb{P} \left( d_{k}(f^{(n)}, Y^{(n)}) \geq \delta \right) \leq 2 \exp \left( -\frac{\delta^2 n}{2\sigma^2} \right).
\]
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
\[
\tilde{\tau}_n(\alpha) = \sqrt{-2\sigma^2/n \log(\alpha/2)}.
\]
\end{remark}

\textbf{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.

\begin{theorem}
Assume the regression model (2) and the moment condition (9). Let $f^{(n)} \in X_k$ be such that $s_{k-1}(f^{(n)}) \geq \epsilon$. Then
\[
\mathbb{P} \left( k_{\epsilon/2}(Y^{(n)}) = k \right) \geq 1 - 2 \exp \left( -\frac{\epsilon^2 n}{8\nu + 4k\epsilon} \right).
\]
\end{theorem}

\textbf{Proof} First suppose that $k > 0$. Notice that by (12) 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
\[
\mathbb{P} \left( k_{\epsilon/2}(Y^{(n)}) \neq k \right) \leq \mathbb{P} \left( \max_{j \not\in \mathcal{I}_0} \left| s_j(Y^{(n)}) - s_j(f^{(n)}) \right| \geq \frac{\epsilon}{2} \right) \leq 2 \exp \left( -\frac{\epsilon^2 n}{8\nu + 4k\epsilon} \right).
\]

We stress that the bound in Theorem 5 is remarkably simple, as it depends on the signatures $s_j(f^{(n)})$, $j \in \{0, \ldots, k-1\}$, 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 $k$.  

16
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:

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

(17)

for some \(\epsilon > 0\) and for some \(j \in \{0, \ldots, n-1\}\) that is a priori not known. Note that there exists no statistical testing procedure that can asymptotically (as the number of observation \(n \to \infty\)) detect signals with intensity as in (17) for \(\epsilon < 0\) with positive detection power, see, e.g., [13]. For \(\epsilon > 0\), 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 (17) for any \(\epsilon > 0\):

**Theorem 6 (Kolmogorov signatures and sparse signals)** Let \(f_n : [0, 1] \to \mathbb{R}\) be as in (17), and let \(Y_i = f_n(x_i) + \epsilon_i\), where \(\epsilon_1, \ldots, \epsilon_n\) \(\text{i.i.d.} \sim \mathcal{N}(0, 1)\). Then for any \(\alpha \in (0, 1)\) one has

\[
\lim_{n \to \infty} \mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) = 1) = 0,
\]

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

**Proof** We have

\[
\mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) = 1) \leq \mathbb{P}(k_{\tau_n(\alpha)}(Y^{(n)}) \geq 1) = \mathbb{P}(d_0(Y^{(n)}) \geq \tau_n(\alpha)) \leq \mathbb{P}(d_K(Y^{(n)}, 0) \geq \tau_n(\alpha)),
\]

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

\[
\mathbb{P}(d_K(Y^{(n)}, 0) \geq \tau_n(\alpha)) \leq \mathbb{P} \left( \max_{n=1, \ldots, n} \sum_{i=1}^{m} \epsilon_i + \mu_n \geq \tau_n(\alpha) \right) 
\leq \mathbb{P} \left( \max_{n=1, \ldots, n} \sum_{i=1}^{m} \epsilon_i + \frac{\mu_n}{\sqrt{n}} \geq \tau_n(\alpha) \right) 
= \mathbb{P} \left( \max_{n=1, \ldots, n} \sum_{i=1}^{m} \epsilon_i + \frac{\mu_n}{\sqrt{n}} \geq \sqrt{n} \tau_n(\alpha) \right).
\]

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

\[
\frac{1}{\sqrt{n}} \max_{n=1, \ldots, n} \sum_{i=1}^{m} \epsilon_i \xrightarrow{\mathcal{D}} \sup_{0 \leq x \leq 1} B(x),
\]

where \(B\) denotes a standard Brownian motion on \([0, 1]\). \(\square\)
4.2 Simulations using Kolmogorov signatures

We illustrate the validity of our approach by means of a simulation study for the signals \textit{blocks} and \textit{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 \textit{blocks} and $\sigma = \sqrt{n}/256$ for \textit{bumps} and computed the average value of $\Delta(Y)$ in 1000 Monte-Carlo simulations. The results in Fig. 4 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 \textit{blocks} and \textit{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 \textit{blocks} and \textit{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 Sup norm based 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

$$f_n(x) = \begin{cases} 
\delta_n & \text{if } x \in [1/3, 2/3), \\
0 & \text{otherwise,}
\end{cases} \quad (18)$$

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

$$\delta_n \sqrt{n} \to \infty, \quad (19)$$

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 $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and let $f_n : [0, 1] \to \mathbb{R}$ be as in (18) with $\delta_n \sqrt{\log n} \to 0$. For any arbitrary sequence $q_n \in \mathbb{R}$

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

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

$$\lim_{n \to \infty} \mathbb{P}\left(\max_{1 \leq i \leq n} Z_i \leq a_n + b_n \epsilon\right) = e^{-e^{-x}}. \quad (20)$$

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

$$a_n = \sqrt{2 \log n} - \left(1/2 \log \log n + \log 2 \sqrt{n} / \sqrt{2 \log n}\right), \quad b_n = 1 / \sqrt{2 \log n}. \quad (20)$$

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

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

$$\Delta_m = \min_{h \in \mathbb{R}^{2m}, h_1 \leq h_2 \leq \cdots \leq h_{2m}} \|\epsilon - h\|_\infty. \quad (20)$$

Then, with $a_m$ and $b_m$ as in (20),

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

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

$$\|\epsilon - h\|_\infty \geq \max \left\{ M^{(1)} - h_m, h_m - M_{(2)} \right\}. \quad (20)$$
Hence,
\[ \Delta_m \geq \min_{\xi \in \mathbb{R}} \max \{ M^{(1)} - \xi, \xi - M^{(2)} \} = \frac{1}{2} (M^{(1)} - M^{(2)}) \stackrel{d}{=} M^{(1)}, \]
where \( A \stackrel{d}{=} B \) means that \( A \) and \( B \) are equally distributed. This implies that
\[
\lim_{m \to \infty} P(\Delta_m \leq a_m + b_m x) \leq \lim_{m \to \infty} P(M^{(1)} \leq a_m + b_m x) = e^{-e^{-x}},
\]
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
\[
P\left( k_{q_n}^\ast(Y^{(n)}) \geq 1 \right) = P\left( s_{0,\infty}(Y^{(n)}) \geq q_n \right) \quad \text{and} \quad P\left( k_{q_n}^\ast(Y^{(n)}) > 1 \right) = P\left( s_{1,\infty}(Y^{(n)}) \geq q_n \right). \tag{22}
\]
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
\[
P\left( k_{q_n}^\ast(Y^{(n)}) \geq 1 \right) \leq P\left( d_\infty(f_n, Y) \geq q_n - \delta_n/2 \right) \tag{23}
\]
\[
= P\left( d_\infty(f_n, Y) - a_n \geq \frac{q_n - a_n}{b_n} - \frac{\delta_n}{2b_n} \right)
\]
\[
= P\left( d_\infty(f_n, Y) - a_n \geq \frac{q_n - a_n}{b_n} + o(1) \right),
\]
with \( a_n \) and \( b_n \) as in (20). Since \( d_\infty(f_n, Y) = \max_{i=1,\ldots,n} |e_i| \) it follows that for any \( x \in \mathbb{R} \) one has
\[
P\left( d_\infty(f_n, Y) \geq x \right) \leq 2P\left( \max_{i=1,\ldots,n} e_i \geq x \right) \text{ by symmetry. Therefore,}
\]
\[
\lim_{n \to \infty} P\left( d_\infty(f_n, Y) - a_n \geq x \right) \leq 2\left( 1 - e^{-e^{-x}} \right). \tag{24}
\]
Further, for \( i = 0, \ldots, 5 \) we define
\[ \Delta_i^+ = \min_{h \in \text{Ham}} \| \pm (Y_{im+1}, \ldots, Y_{(i+1)m}) - h \|_\infty. \]
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 \( s_{1,\infty}(Y^{(n)}) \geq \min \{ \Delta_0^-, \Delta_0^+, \Delta_1^-, \ldots, \Delta_5^- \} \). Note that \( \Delta_0^-, \Delta_0^+, \ldots, \Delta_5^- , \Delta_5^+ \) are identically distributed and independent asymptotically. Therefore,
\[
P\left( k_{q_n}^\ast(Y^{(n)}) > 1 \right) = \lim_{n \to \infty} P\left( s_{1,\infty}(Y^{(n)}) \geq q_n \right) \geq \lim_{n \to \infty} P\left( \min \{ \Delta_0^-, \Delta_0^+, \ldots, \Delta_5^- \} \geq q_n \right)
\]
\[
= \lim_{n \to \infty} \left( 1 - P\left( \Delta_0^- < q_n \right) \right)^{12}.
\]
In order to prove the assertion, we show that for some \( \beta \in (0, 1) \)
\[
\lim_{n \to \infty} P\left( k_{q_n}^\ast(Y^{(n)}) \geq 1 \right) \geq 1 - \beta
\]
20
already implies
\[
\lim_{n \to \infty} \mathbb{P}(k_{q_n}^{\infty}(Y(n)) > 1) > 0
\]
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 (23) and (24) shows that
\[
\lim_{n \to \infty} \mathbb{P}(k_{q_n}^{\infty}(Y(n)) \geq 1) \geq 1 - \beta
\]
implies
\[
q_n \leq a_n + b_n z_\beta + o(b_n)
\]
where \(z_\beta\) 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
\[
\lim_{n \to \infty} \mathbb{P}(k_{q_n}^{\infty}(Y(n)) \geq 1) \geq 1 - \exp(-e^{-e^{-z_\beta}}) - \frac{\log 6}{12}.
\]
Here the last inequality follows from Lemma 6 together with \(\frac{b_n}{b_m} \to 1\) and \(\frac{a_n - a_m}{b_m} \to \log 6\). The proof is then completed by observing that \(z_\beta + \log 6 < \infty\), which yields \(1 - e^{-e^{-z_\beta}} > 0\).

\[\blacksquare\]

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 \(d_\infty\)-ball \(D_\alpha(F)\) of radius \(\alpha \geq 0\) around \(F\). We refer to \(D_\alpha(F)\) as the \(\alpha\)-tube around \(F\). The taut string, denoted by \(F_\alpha\), 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
\[
F_\alpha(0) = F(0) \quad \text{and} \quad F_\alpha(1) = F(1).
\]

For existence and uniqueness, we refer to [20, 21]. \(F_\alpha\) is Lipschitz continuous for all \(\alpha > 0\) (see [20], proof of Lemma 2); thus its derivative \(f_\alpha\) (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 \(F_\alpha\) 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 \(F_\alpha\) minimizes the number of modes among all function \(g \in \mathcal{L}\) with \(d_K(f, g) \leq \alpha\). \(\blacksquare\)
The proof requires some preparation. Let the top and bottom functions of the $\alpha$-tube around the antiderivative $F$ of $f \in L$ 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 $\alpha$-tube.

**Lemma 7 (Grasmair and Obereder [21])** For every $\alpha > 0$, the taut string $F_\alpha$ 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, $F_\alpha$ is piecewise affine outside of $S_{B,\alpha} \cup S_{T,\alpha}$. □

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 $f_\alpha$ corresponds to a point (or interval) where its antiderivative $F_\alpha$ 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 $F_\alpha$ 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 $F_\alpha$ is convex (concave) on $I$ but not on any interval that properly contains $I$. □
Observe that by Lemma 7, if $F_\alpha$ 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 $F_\alpha$ if $I$ is a maximally affine interval of $F_\alpha$ and $F_\alpha$ is convex (concave) on some non empty neighborhood of $a$ and concave (convex) on some non empty neighborhood of $b$.

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 $F_\alpha$, 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 $F_\alpha(b) = T_\alpha(a)$. □

Moreover we have:

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

(i) The number of maximally convex, the number of maximally concave, and the number of inflection intervals of $F_\alpha$ 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. □

**Proof** Let $T_\alpha$ and $B_\alpha$ denote the top and bottom of the $\alpha$-tube around $F_\alpha$, respectively. Since $F_\alpha$ is continuous, the graphs of $T_\alpha$ and $B_\alpha$ are compact sets. Let $I$ be a maximally convex, a maximally concave, or an inflection interval of $F_\alpha$. By Definitions 3 and 4 and Lemma 7, the graph of $F_\alpha$ restricted to $I$ must then contain an affine segment that connects $T_\alpha$ with $B_\alpha$ (or $B_\alpha$ with $T_\alpha$). Therefore, the arc length of the graph of $F_\alpha$, restricted to $I$, is bounded from below by the Euclidean distance $d_\alpha$ between the graphs of $B_\alpha$ and $T_\alpha$. Since these sets are compact and disjoint, one has $d_\alpha > 0$. Since $d_\alpha$ is independent of $I$, and since $F_\alpha$ is Lipschitz, it follows that the length of $I$ is bounded from below by a number that only depends on $\alpha$ and the Lipschitz constant of $F_\alpha$. 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

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

and $f_\alpha(t) = \lim_{s \to t} f_\alpha(s)$ for $t \in \{0, 1\}$. Then the number of positive inflection intervals of $F_\alpha$ equals the number of modes of $f_\alpha$, and this number is finite. □
Proof First notice that the definition of \( f_\alpha(0) \) and \( f_\alpha(1) \) is meaningful since \( F_\alpha \) is affine in some neighborhood of 0 and 1.

If \( F_\alpha \) 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 \( P = [t_0, \ldots, t_{|P|}] \) of \([0, 1]\). Notice that \( f_\alpha \) is nowhere decreasing (nowhere increasing) on intervals where \( F_\alpha \) is convex (concave). Hence, for \( t_i \) to count a mode of \( f_\alpha \), 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 \( F_\alpha \). Since, by assumption, \( F_\alpha \) 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 \( f_\alpha \) 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 \( \mathcal{I}^+(f_\alpha) \). It follows that \( \mathcal{I}^+[F_\alpha] \geq \mathcal{M}(f_\alpha) \).

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 \( \mathcal{I}^+[F_\alpha] \leq \mathcal{M}(f_\alpha) \).

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 \( F_\alpha \). By Lemma 8, \( F_\alpha(a) = T_\alpha(a) \), \( F_\alpha(b) = B_\alpha(b) \), and \( F_\alpha \) 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 \( g : [a, b] \to \mathbb{R} \) 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 \( F_\alpha \) is not empty. Therefore, one can choose a partition \( P = [t_0, \ldots, t_{|P|}] \) 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 \( F_\alpha \) 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 \( f_\alpha \) precisely for every positive inflection interval of \( F_\alpha \). 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 \( f_\alpha \) of the antiderivative \( F_\alpha \) for all \( \alpha \geq 0 \). For example, let \( X \) be the space of piecewise constant functions. For \( \alpha \) large enough, \( F_\alpha \) is affine on all of \([0,1]\), and its derivative \( f_\alpha \) has no modes. If \( f \) has any modes at all, then by lowering \( \alpha \) continuously, \( F_\alpha \) will at some point develop a positive inflection interval below some threshold \( \alpha_0 > 0 \). By Theorem 8 and Lemma 10, the value of \( \alpha_0 \) 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 \( \alpha_k \) as the smallest \( \alpha \) for which \( f_\alpha \) 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 \( \alpha \) (in increasing order) at which the number of modes of \( f_\alpha \) decreases. To this end, we work with the space \( X \) of piecewise constant functions on a fixed partition \( 0 = t_0 < t_1 < \cdots < t_n = 1 \) of \([0,1]\). Notice that since we require \( X \subset \mathcal{L} \), we have \( f(t_i) = \frac{1}{2}(f(t_{i,1,4}) + f(t_{i,5,1})) \) 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 \( F_\alpha \) 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 \( F_\alpha(t) = F(t) + \alpha \) (\( F_\alpha(t) = F(t) - \alpha \)). \( \square \)

Fix \( \alpha \geq 0 \). Let \( I = (a, b) \subseteq (0, 1) \) be an open interval, and let \( f_\alpha \) be constant on \((a, b)\). We call \( I \) regular for \( f_\alpha \) 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 \( f_\alpha \) 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 \( f_\alpha \) if either \( a = 0 \) and \( b < 1 \) or \( a > 0 \) and \( b = 1 \), and \((a, b)\) is the largest such interval on which \( f_\alpha \) is constant. As a consequence of Lemma 11 we obtain:

Corollary 2 Away from discontinuities, \( f_\alpha \) has the following form: either

- \( t \) lies on a regular interval \( I = (a, b) \) of \( f_\alpha \) with value \( f_\alpha(t) = \frac{F(b) - F(a)}{b - a} \),

- \( t \) lies on a locally minimal/maximal interval \( I = (a, b) \) of \( f_\alpha \) with value \( f_\alpha(t) = \frac{F(b) - F(a) + 2\alpha}{b - a} \), or

- \( t \) lies on a boundary interval of \( f_\alpha \) with value \( f_\alpha(t) = \frac{F(b) - F(a) + \alpha}{b - a} \). \( \square \)

This corollary is central for our computation of Kolmogorov signatures. First observe that a value of a maximal interval is continuously decreasing with growing \( \alpha \), the value of a minimal interval is continuously increasing, and the value of a regular interval remains unchanged. Moreover, if \( \alpha \) is increased only slightly, then the discontinuities of \( f_\alpha \) 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 \( f_\delta \) coincide with those of \( f_\alpha \) for all \( \beta \) with \( \alpha \leq \beta < \alpha + \delta \). Moreover, if \( t \) lies on a regular interval of \( f_\alpha \), then \( f_\beta(t) = f_\alpha(t) \). \( \square \)
Proof Define $G_\beta$ by the properties of Lemma 11, using the discontinuities of $f_\alpha$, i.e., if $t$ is an increasing (decreasing) discontinuity of $f_\alpha$, 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 $\delta$ sufficiently small, the discontinuities of $g_\beta = G'_\beta$ have the same type as those of $f_\alpha$. But since $F_\beta$ is uniquely defined by the properties of Lemma 11 with respect to these discontinuities, we must have $F_\beta = G_\beta$.

As a consequence, for every $\alpha \geq 0$, there exists a minimal number $\mu(\alpha) > \alpha$ such that $f_\beta$ and $f_\alpha$ have the same points of discontinuity for all $\beta$ with $\alpha \leq \beta < \mu(\alpha)$ but the set of points of discontinuity of $f_{\mu(\alpha)}$ is different from that of $f_\alpha$. We call $\mu(\alpha)$ the merge value of $\alpha$. The merge value is the smallest number strictly greater than $\alpha$ 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 $f_\alpha$ 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 $f_\alpha$. 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 $f_\alpha$, 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

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

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

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

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

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

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

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

We define the sequence of merge values $\mu_1 < \mu_2 < \mu_3 < \ldots$ 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 $f_\alpha$ decreases with increasing $\alpha$.

Observe that the merge value candidates of $f_{\mu_{i+1}}$ are equal to those of $f_\mu$ 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 $f_{\mu_i}$. 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 $\mu_i$. 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 $f_\alpha$ has decreased upon a merge, the value $\alpha$ 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:             recomputes 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] S. Balakrishnan, A. Rinaldo, D. Sheehy, A. Singh, and L. A. Wasserman. Minimax rates for homology inference. Journal of Machine Learning Research - Proceedings Track, 22: 64–72, 2012.

[2] U. Bauer, C. Lange, and M. Wardetzky. Optimal topological simplification of discrete functions on surfaces. Discrete & Computational Geometry, 47(2):347–377, 2012.

[3] P. Bendich, T. Galkovskyi, and J. Harer. Improving homology estimates with random walks. Inverse Problems, 27(12):124002+, 2011.
[4] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., second edition, 1999. A Wiley-Interscience Publication.

[5] P. Bubenik and P. T. Kim. A statistical approach to persistent homology. *Homology, Homotopy and Applications*, 9(2):337–362, 2007.

[6] P. Bubenik, G. Carlsson, P. T. Kim, and Z.-M. Luo. Statistical topology via Morse theory persistence and nonparametric estimation. In M. A. G. Viana and H. P. Wynn, editors, *Algebraic Methods in Statistics and Probability II*, volume 516 of *Contemporary Mathematics*, pages 75–92. American Mathematical Society, 2010.

[7] H. P. Chan and G. Walther. Detection with the scan and the average likelihood ratio. *arXiv:1107.4344v1*, 2013.

[8] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. *Computer Graphics Forum*, 28(5):1393–1403, 2009.

[9] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.

[10] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

[11] P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *The Annals of Statistics*, 29(1):1–65, 2001. With discussion and rejoinder by the authors.

[12] P. L. Davies and A. Kovac. Densities, spectral densities and modality. *The Annals of Statistics*, 32(3):1093–1136, 2004.

[13] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.

[14] D. L. Donoho. One-sided inference about functionals of a density. *The Annals of Statistics*, 16(4):1390–1420, 1988.

[15] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2):301–369, 1995. With discussion and a reply by the authors.

[16] H. Edelsbrunner and J. L. Harer. *Computational Topology: An Introduction*. AMS, 2010.

[17] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

[18] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 2014.

[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(369):pp. 42–56, 1980.
[20] M. Grasmair. The equivalence of the taut string algorithm and BV-regularization. *Journal of Mathematical Imaging and Vision*, 27(1):59–66, 2007.

[21] M. Grasmair and A. Obereder. Generalizations of the taut string method. *Numerical Functional Analysis and Optimization*, 29(3-4):346–361, 2008.

[22] J. A. Hartigan. Testing for antimodes. In Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization, pages 169–181. Springer Berlin Heidelberg, 2000.

[23] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The Annals of Statistics*, 13(1):pp. 70–84, 1985.

[24] Y. Ingster and I. Suslina. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, volume 169 of *Lecture Notes in Statistics*. Springer, 2003.

[25] J. Kloke and G. Carlsson. Topological De-Noising: Strengthening the topological signal, 2010. arXiv:0910.5947.

[26] E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25(1):387–413, 1997.

[27] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, 2000.

[28] D. R. Sheehy. A multicover nerve for geometric inference. In *CCCG: Canadian Conference in Computational Geometry*, 2012.

[29] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*, volume 59. Siam, 2009.

[30] B. W. Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):pp. 97–99, 1981.

[31] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

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