

Novelty detection on path space

Ioannis Gasteratos

*Institute of Mathematics
TU Berlin*

I.GASTERATOS@TU-BERLIN.DE

Antoine Jacquier

*Department of Mathematics
Imperial College London*

A.JACQUIER@IMPERIAL.AC.UK

Maud Lemerrier

*Mathematical Institute
University of Oxford*

MAUD.LEMERCIER@MATHS.OX.AC.UK

Terry Lyons

*Mathematical Institute
University of Oxford*

TERRY.LYONS@MATHS.OX.AC.UK

Cristopher Salvi

*Department of Mathematics
Imperial College London*

C.SALVI@IMPERIAL.AC.UK

Abstract

We frame novelty detection on path space as a hypothesis testing problem with signature-based test statistics. Using transportation-cost inequalities of Gasteratos and Jacquier (2023), we obtain tail bounds for false positive rates that extend beyond Gaussian measures to laws of RDE solutions with smooth bounded vector fields, yielding estimates of quantiles and p-values. Exploiting the shuffle product, we derive exact formulae for smooth surrogates of conditional value-at-risk (CVaR) in terms of expected signatures, leading to new one-class SVM algorithms optimising smooth CVaR objectives. We then establish lower bounds on type-II error for alternatives with finite first moment, giving general power bounds when the reference measure and the alternative are absolutely continuous with respect to each other. Finally, we evaluate numerically the type-I error and statistical power of signature-based test statistic, using synthetic anomalous diffusion data and real-world molecular biology data.

Keywords: Anomaly detection, hypothesis testing, signature features, rough paths

MSC 2020: 60L10, 60L20, 62M07

Contents

1	Introduction	2
1.1	Contributions	4
1.2	Related work	5
2	Test statistics on path space	6
2.1	One-class support vector machines	7
2.2	Distance to the expected signature	8
2.3	Conformance score	9

3	Hypothesis testing for novelty detection on path space	9
3.1	Sample-free quantile estimates via expected signature	10
3.2	Probabilistic quantile estimates via transportation-cost inequalities	12
4	Experiments	18
4.1	Anomalous diffusions	19
4.2	Anomaly detection in RNA direct sequencing data	21
5	Conclusion	22
A	Signature test statistics	23
A.1	Conformance score	23
A.2	One-class support vector machines	25
B	Conformance score for non-Gaussian streamed data	26
C	Experimental Details	30

1 Introduction

Determining whether an observed trajectory is consistent with a reference process is a fundamental novelty detection problem, with applications ranging from detecting unusual particle motion in biological cells (Muñoz Gil et al., 2021) to identifying rare astronomical signals in telescope data (Wilensky et al., 2019) and detecting malicious activities in network traffic (Cochrane et al., 2021). Most existing approaches cast this as one-class classification on time series data and focus primarily on feature engineering or representation learning to obtain suitable embeddings (Li and Jung, 2023; Zamanzadeh Darban et al., 2024). In this paper, we frame trajectory-based novelty detection as a hypothesis testing problem on path space. The goal is to test whether an observed path $X : [0, T] \rightarrow \mathbb{R}^d$ is drawn from a reference measure μ representing normal behaviour, or from an alternative distribution $\nu \neq \mu$, i.e. we want to test the null hypothesis

$$H_0 : X \sim \mu \text{ against } H_1 : X \sim \nu \neq \mu,$$

where ν and μ are two distributions on some space of paths $\mathcal{X} \subset C([0, T]; \mathbb{R}^d)$, $T > 0$. To perform this hypothesis test, we map each sample path X to a scalar statistic via a map $f : \mathcal{X} \rightarrow \mathbb{R}$ which takes the form $f(X) = \langle w, S(X) \rangle$ where $w \in T((\mathbb{R}^d)^*)$ is a linear functional acting on the signature of the process, given by the formal tensor series of iterated integrals

$$S(X) = 1 + \sum_{k=1}^{\infty} S^k(X) \in T((\mathbb{R}^d)), \text{ with } S^k(X) = \int_{0 < t_1 < \dots < t_k < T} dX_{t_1} \otimes \dots \otimes dX_{t_k},$$

and $\langle \cdot, \cdot \rangle$ denotes the natural pairing between the infinite formal series $T((\mathbb{R}^d)) = \prod_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k}$ and the finite sequences in $T((\mathbb{R}^d)^*) = \bigoplus_{k=0}^{\infty} ((\mathbb{R}^d)^{\otimes k})^*$. The null hypothesis H_0 is rejected if the observed path X belongs to the rejection region Ω_r of the form

$$X \in \Omega_r \iff \langle w, S(X) \rangle > r. \quad (1.1)$$

This framework entails two types of errors. If $X \sim \mu$, but H_0 is rejected based on (1.1), the test incurs a type-I error. The probability of making such an error—the false positive rate—is given by $\alpha := \mu(x \in \Omega_r)$. Conversely, if $X \sim \nu \neq \mu$, but H_0 is not rejected based on (1.1), a type-II error occurs, with probability $\beta := \nu(x \notin \Omega_r)$. The power of our test against the alternative ν is given by $1 - \beta$.

The central challenge is to choose the linear functional w and the threshold r to achieve high power against a broad class of alternative distributions ν on \mathcal{X} , while maintaining a fixed significance level $\alpha \in [0, 1]$ (i.e., controlling the probability of type-I error). In statistical hypothesis testing, the test can equivalently be formulated in terms of p-values. Specifically, it is common to report the probability of obtaining test results at least as extreme as the result actually observed, as a measure of significance of the test (here, evidence of path X being a novelty) rather than whether or not the null hypothesis was rejected at a pre-determined level of significance α .

Regarding critical value selection, the threshold r corresponding to a given significance level α is determined by $r(\alpha) = Q_{f(X)}(1 - \alpha)$, the $(1 - \alpha)$ -quantile of the null distribution of the test statistic $f(X)$. The α -quantile of a real-valued random variable Z with law P_Z is defined as

$$Q_Z(\alpha) := \inf\{r \in \mathbb{R} \mid P_Z(z \leq r) \geq \alpha\}.$$

In risk management contexts, quantiles are also known as values at risk. In the sequel they will be denoted by $\text{VaR}_\alpha(Z)$. If the cumulative distribution function F of the random variable $f(X)$ under the null hypothesis is continuous, this reduces to the relation $r(\alpha) = F^{-1}(1 - \alpha)$. However, in general, for a given measure μ on \mathcal{X} and test statistic $f : \mathcal{X} \rightarrow \mathbb{R}$ the distribution of $f(X)$ is not available in closed form. Therefore, it is rarely possible to conduct the above test. A common workaround is to use a calibration sample $X^{(1)}, \dots, X^{(n)} \stackrel{\text{i.i.d.}}{\sim} \mu$ and the corresponding empirical estimator $F_n(f(x)) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{f(X^{(i)}) \leq f(x)\}$ to determine the critical value of the test. While simple and widely used, this approach comes with important caveats (Bates et al., 2023). Conditioned on a fixed calibration set, the false positive rate may exceed the target level α ; it is only guaranteed to be controlled marginally over calibration sets. Moreover, when empirical p-values are used, the dependencies introduced by sharing the same calibration set can invalidate certain multiple testing procedures, such as Fisher’s combination test. These limitations are especially relevant in novelty detection applications on data streams, where practitioners must often test multiple data segments sequentially. For instance, in radio astronomy, the task of identifying radio-frequency interference (RFI) in telescope visibility measurements involves repeatedly testing thousands of short time-frequency windows for anomalies (Arrubarrena et al., 2024). A similar challenge arises in biological signal analysis, where electrical readouts from RNA molecules are scanned segment by segment to identify potential chemical modifications (Leger et al., 2021). A solution is to derive an upper bound on the tail probabilities of the test statistic, ensuring that the rejection threshold controls the false positive rate regardless of the calibration sample, or with high probability over the sample.

Regarding the choice of score function (one-class classifier) $f : \mathcal{X} \rightarrow \mathbb{R}$, it is important to ensure that the test has high statistical power, i.e. is effective at flagging outliers. Since the signature has the *universal approximation property* on compact subsets of \mathcal{X}

(the set of functions of the form $f(x) = \langle w, S(x) \rangle$ is dense in the set of continuous functions on any such compact subset), the class of test statistics we consider is extremely rich (Arribas, 2018). Leveraging this universal approximation property and other algebraic features, signature methods have rapidly gained traction across domains such as quantitative finance (Perez Arribas, 2020; Cuchiero and Möller, 2023; Abi Jaber and Gérard, 2025; Cuchiero and Möller, 2023; Bonesini et al., 2024), cybersecurity (Cochrane et al., 2021), information theory (Salvi, 2021; Salvi et al., 2023; Shmelev and Salvi, 2024), and quantum computing (Crew et al., 2025). They underpin universality results for *neural differential equations* (Morrill et al., 2021; Arribas et al., 2020) and *state-space models* (SSMs) (Muça Cirone et al., 2024; Muça Cirone and Salvi, 2025a; Walker et al., 2025). In generative modelling, signatures have been used to synthesise financial time series (Buehler et al., 2020), act as universal nonlinearities in Seq2Seq models (Kidger et al., 2019), and provide representation spaces for training *score-based diffusion models* (Barancikova et al., 2025). An introduction to signature methods in machine learning can be found in (Cass and Salvi, 2024) and a survey of recent applications in (Fermanian et al., 2023). Various Python packages for signature computations are readily available, such as `esig` (Lyons and al, 2010), `iisignature` (Reizenstein and Graham, 2018), `signatory` (Kidger and Lyons, 2021) and more recently `pySigLib` (Shmelev and Salvi, 2025). In this paper we made use of `pySigLib` because of its superior performance compared to the other alternatives.

1.1 Contributions

We leverage transportation-cost inequalities recently established by Gasteratos and Jacquier (2023) to derive an upper bound on the false positive rate $P_I(r) := \mu(f(x) > r)$ for signature-based test statistics on path space. These tail estimates extend beyond Gaussian measures to a much broader class of reference measures μ , including laws of solutions of RDEs with Gaussian drivers with bounded and sufficiently smooth vector fields. Our tail estimates yield theoretical upper bounds on the critical value $r(\alpha)$ for a desired level of significance α . When numerically tractable, these provide valid (or super-uniform) p-values for hypothesis tests, though potentially being more conservative and decreasing the power.

Beyond hypothesis testing, extreme events and tails of random variables are central in problems of risk management. For example, in finance, quantiles of loss distributions serve as value-at-risk (VaR) measures, while the conditional value-at-risk (CVaR) at a given confidence level captures the expected loss given that the loss exceeds the VaR at that level. CVaR admits a dual formulation as the solution to a minimisation problem involving the max-function, which is non-differentiable. In this paper, we consider smooth surrogates where the max-function is replaced by a polynomial. Leveraging the shuffle product property of the signature, we show that for any such smooth surrogate, when $f(X) = \langle w, S(X) \rangle$ is our signature-based test statistic for the measure μ , the resulting smooth CVaR admits an exact analytic expression in terms of the expected signature of μ , whenever $\mathbb{E}_\mu[|f(X)|] < \infty$. This yields new one-class support vector machine algorithms for novelty detection that directly optimise smooth CVaR objectives over the space of signature-based statistics.

We derive lower bounds on the type-II error probability $P_{II}(r) = \nu(f(x) \geq r)$, for any alternative ν that has finite first moment. These provide informative upper bounds on the power of the test against any alternative that has finite relative entropy $\mathbf{H}(\nu|\mu)$. While the

transportation-cost inequalities from Gasteratos and Jacquier (2023) imply well-known tail estimates (Cass et al., 2013) and the type-I error bound can be retrieved using these prior results for a specific class of reference measures μ , the lower bound on the type-II error fundamentally requires the new results in (Gasteratos and Jacquier, 2023).

1.2 Related work

There exists a large body of machine learning methods for anomaly detection on time-series data. One prominent line of work has focused on extending well-established *algorithms* for anomaly detection on vectorial data via functional or signature embeddings. For instance, Functional Isolation Forest (IF) (Staerman et al., 2019) adapts Isolation Forest to function spaces, and Signature IF (Staerman et al., 2024) further leverages rough path signatures to handle multivariate processes; nearest neighbours based methods using signature-derived features have been proposed (Shao et al., 2020) and One-class SVM has been equipped with multiscale signature features (Mignot et al., 2024). Another direction leverages the representation learning capabilities of neural networks. Notably, Ruff et al. (2018, 2019) combines deep learning with Support Vector Data Description (SVDD), where neural networks learn feature representations from high-dimensional data such as images or text.

On the theoretical side, Cass et al. (2024) derives the distribution of Mahalanobis distances under Gaussian measures on Hilbert spaces, but a formal hypothesis testing procedure for anomaly detection on path space—especially under non-Gaussian signature embeddings—remains absent. Consequently, most approaches are evaluated using threshold-independent metrics such as the Area Under the Receiver Operating Characteristic curve (AUROC), which measure the ability of a score function to rank anomalies above Normal observations when labelled data is available. While useful for benchmarking, such metrics do not yield a principled framework for making and reporting discoveries at a controlled false positive rate. In this work, we close this gap by establishing explicit bounds on the type-I error for anomaly detection on path space under weaker distributional assumptions.

A related line of work by Bates et al. (2023) establishes confidence bounds on the false positive rate. These bounds hold with high probability over the random draw of the calibration set, providing stronger guarantees on the control of the false positive rate compared to the vanilla conformal inference approach. In contrast, the bounds we establish are not confidence intervals but deterministic tail bounds: they hold with full probability and guarantee type-I error control uniformly across thresholds. Moreover, the tightness of the upper confidence bound from Bates et al. (2023) which is of the form $h(F_n(r))$ where $F_n(r) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[f(X^{(i)}) > r]$ is limited by the size of the calibration set.

Our results hold for probability measures satisfying suitable generalised Transportation-Cost Inequalities (TCIs) from Gasteratos and Jacquier (2023), which imply well-known estimates for the p -variation of the law of solutions of RDEs driven by Gaussian rough paths (Cass et al., 2013). Determining the tail behaviour of a variable of interest (here a test statistic for anomaly detection), i.e. the rate at which the probability of exceeding increasingly high values decreases, is a central question in extreme value theory (EVT). However, in this field, the goal is to describe the statistics of observed extremes using asymptotic approximations. The number of available observations strongly affects the esti-

mation of the shape parameter of the general distributions used in EVT, which ultimately determines the limiting type. Importantly, the Weibull distribution should not be confused with the third limiting type of the EVT, termed the *reversed Weibull*.

Another hypothesis testing problem that has been studied on path space is the two-sample test (Wynne and Duncan, 2022). Chevyrev and Oberhauser (2022) introduced a signature-based maximum mean discrepancy (MMD) framework for comparing path distributions, which is used for example in Andr  s et al. (2024) to compare sample paths of fractional Brownian motion. The two-sample setting tests whether two unknown distributions are equal based on samples from each. In this paper, we consider the one-sample goodness-of-fit problem testing whether a single observed path originates from a fixed reference distribution μ . Additionally, while the test may be conducted in practice using samples from μ , our theoretical framework provides new tail and CVaR estimates for the population-level measure μ .

2 Test statistics on path space

We provide a brief overview of commonly used anomaly scores on path space, which are given (or can be approximated) by a linear functional of the signature. These include distance-based score functions as well as solutions of one-class SVM optimisation problems. More details can be found in Section A.

In the sequel we fix $T > 0$ and consider bounded variation paths $C^{1\text{-var}}([0, T]; \mathbb{R}^d)$, namely continuous paths $x : [0, T] \rightarrow \mathbb{R}^d$ with bounded total variation on each sub-interval $[s, t] \subset [0, T]$, i.e.

$$\|x\|_{1\text{-var}, [s, t]} := \sup_D \sum_k \|x_{t_{k+1}} - x_{t_k}\| < \infty,$$

where $\|\cdot\|$ is the Euclidean norm and where the supremum is taken over all finite dissections $D = \{s = t_1 < \dots < t_K = t\}$. We consider test statistics that are linear combination of signature coordinates of the path. The coordinate associated with a word $(i_1, \dots, i_k) \in \{1, \dots, d\}^k$ of length k is

$$S^{i_1, \dots, i_k}(x) = \int_{0 < t_1 < \dots < t_k < T} dx_{t_1}^{i_1} \dots dx_{t_k}^{i_k},$$

where the integral is to be understood as the Riemann-Stieltjes integral. While the signature is an infinite formal series

$$S(x) = 1 + \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k} S^{i_1, \dots, i_k}(x) e_{i_1} \dots e_{i_k} \in T((\mathbb{R}^d)) = \prod_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k},$$

where $\{e_i : i = 1, \dots, d\}$ is a basis of \mathbb{R}^d , the truncated signature at level $N \geq 1$, corresponds to the collection of all signature terms associated to words of length $k \leq N$:

$$S_N(x) = 1 + \sum_{k=1}^N \sum_{i_1, \dots, i_k} S^{i_1, \dots, i_k}(x) e_{i_1} \dots e_{i_k} \in T^N(\mathbb{R}^d) = \bigoplus_{k=0}^N (\mathbb{R}^d)^{\otimes k}. \quad (2.1)$$

We consider elements w of $T(\mathbb{R}^d)^*$, the dual space of the tensor algebra $T(\mathbb{R}^d)$. Products of iterated integrals can be re-expressed as a linear combination of higher-order iterated

integrals. More formally, the shuffle identity states that for any two linear functionals w_1 and w_2 in $T((\mathbb{R}^d))^* \cong T((\mathbb{R}^d)^*)$ (where $T((V))$ is the algebra of formal tensor series on the vector space V), then

$$\langle w_1, S(x) \rangle \langle w_2, S(x) \rangle = \langle w_1 \sqcup w_2, S(x) \rangle, \quad (2.2)$$

where $\sqcup : T(\mathbb{R}^d) \times T(\mathbb{R}^d) \rightarrow T(\mathbb{R}^d)$ is the shuffle product; the reader is referred to (Cass and Salvi, 2024, Section 1.3.3) for more details.

The above may be extended to a broader class of paths, called geometric p -rough path, which are expressed as limits in p -variation of a sequence $(\pi_{[p]}(S(x^n)))$ of truncated signatures of bounded variation paths (x^n) , where $\pi_{[p]}$ is the projection map on $T^{[p]}(\mathbb{R}^d)$. We denote the space of geometric p -rough paths by $G\Omega_p(\mathbb{R}^d)$.

2.1 One-class support vector machines

One-class support vector machines (one-class SVMs) are a widely used class of algorithms for anomaly detection. The central idea is to embed the input space \mathcal{X} into a higher-dimensional space \mathcal{H} , with a Hilbert space structure, via a feature map $\varphi : \mathcal{X} \rightarrow \mathcal{H}$, and then learn a function $g : \mathcal{X} \rightarrow \mathbb{R}$ of the form

$$x \mapsto g(x) := \langle w, \varphi(x) \rangle_{\mathcal{H}} - \rho, \quad (2.3)$$

with $(w, \rho) \in \mathcal{H} \times \mathbb{R}$, that takes positive values within a region of the input space and negative values outside, thereby separating the “normal” data from potential anomalies.

It is well understood that the range of the signature is a subset of a Hilbert space (Salvi et al., 2021a) and the same holds for their truncated versions (Király and Oberhauser, 2019), making them compatible with this framework. More precisely, let $\{\langle \cdot, \cdot \rangle_k : k \in \mathbb{N}\}$ denote a sequence of inner products on the tensor spaces $(\mathbb{R}^d)^{\otimes k}$, where each $\langle \cdot, \cdot \rangle_k$ is the canonical (Hilbert-Schmidt) inner product derived from a fixed inner product on \mathbb{R}^d , with the convention $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$. The subspace

$$\mathcal{H} := \left\{ A \in T((\mathbb{R}^d)) : \|A\|_2 < \infty \right\} \subset T((\mathbb{R}^d)), \quad (2.4)$$

with $\|A\|_2 = \sqrt{\sum_{k=0}^{\infty} \|a^k\|_k^2}$ the ℓ_2 -norm with inner product $\langle A, B \rangle_{\mathcal{H}} = \sum_{k=0}^{\infty} \langle a^k, b^k \rangle_k$, is a Hilbert space. For any space \mathcal{X} of p -rough paths, the signature kernel

$$\kappa(x, y) := \langle S(x), S(y) \rangle_{\mathcal{H}} \in \mathbb{R},$$

is well defined for all $x, y \in \mathcal{X}$. For a truncation level $N \in \mathbb{N}$, we define the truncated inner product $\langle A, B \rangle_{\mathcal{H}_N} := \sum_{k=0}^N \langle a^k, b^k \rangle_k$ and the associated truncated signature kernel

$$\kappa_N(x, y) := \langle S_N(x), S_N(y) \rangle_{\mathcal{H}_N}, \quad (2.5)$$

which corresponds to the kernel induced by the truncated signature map S_N in (2.1). The reader may thus think of φ in (2.3) as being the signature S or its truncated version S_N , with \mathcal{H} and \mathcal{H}_N their respective feature spaces.

In the seminal paper by Schölkopf et al. (2001), the hyperplane parametrised by (w, ρ) is obtained on the basis of n observations $\{x_i\}_{i=1}^n$ in \mathcal{X} by solving the “dual” optimisation problem

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{2} \alpha^\top K \alpha, \quad (2.6)$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{1}{\gamma n}, \quad \text{for all } i = 1, \dots, n \quad e^\top \alpha = 1, \quad (2.7)$$

where the entries of the matrix K are given by $[K]_{i,j} = \langle \varphi(x_i), \varphi(x_j) \rangle_{\mathcal{H}}$. This yields the acceptance region $\Omega_{\text{ocsvm}} := \{x \in \mathcal{X} : \sum_{i=1}^n \alpha_i \kappa(x_i, x) \geq \rho\}$ where $\kappa(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}$ and $\rho = \sum_{j=1}^n \alpha_j \kappa(x_j, x_i)$ for any x_i such that $0 < \alpha_i < \frac{1}{\gamma n}$.

The effectiveness of one-class SVM largely depends on the chosen feature map. For novelty detection on path space with signatures, either truncated at some level $N > 0$, or untruncated, one can use the kernel tricks developed in Király and Oberhauser (2019), Salvi et al. (2021a) and Muça Cirone et al. (2023) for evaluating inner products of signatures. Signature kernels have been applied to hypothesis testing (Salvi et al., 2021b; Lemerrier et al., 2021; Horvath et al., 2023), to causality (Manten et al., 2025), to quantitative finance (Pannier and Salvi, 2024; Muça Cirone and Salvi, 2025b), and have even emerged as infinite-width limits of neural networks (Muça Cirone et al., 2023). They also enable training neural SDEs for time-series generation in fluid dynamics (Salvi et al., 2022), computational neuroscience (Holberg and Salvi, 2024), and again quantitative finance (Issa et al., 2023; Díaz Lozano et al., 2023; Høglund et al., 2023).

After solving (2.6) via sequential minimal optimisation (Chang and Lin, 2011), evaluating the one-class SVM at a new path x requires computing the score $f(x) = \sum_{i=1}^n \alpha_i \kappa(x_i, x)$. In the case where the feature map is the truncated signature, the primal variables $w = \sum_{i=1}^n \alpha_i S_N(x_i)$ may be precomputed and the function

$$f(x) = \langle w, S_N(x) \rangle_{\mathcal{H}_N}$$

may then be evaluated in time complexity $\mathcal{O}(d^N L_x)$, where d is the input path dimension and L_x the number of observations. This amortised evaluation makes one-class SVM particularly attractive when many candidate paths are to be tested against the same measure μ .

2.2 Distance to the expected signature

If μ is a measure on path space with a well-defined expected signature $\mathbb{E}_\mu[S(X)]$, a natural test statistic is obtained by considering the non-negative function $f : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ defined by

$$f(x) := \|S_N(x) - \mathbb{E}_\mu[S_N(X)]\|_2, \quad (2.8)$$

that is, the distance in ℓ_2 -norm between the signature of x , truncated at level N , and the expected truncated signature of the process with law μ . The function in (2.8) can be rewritten as a linear functional of the signature

$$f(x) = \langle w, S(x) \rangle, \quad \text{with } w = \sum_{k=0}^N \sum_{I \in \mathcal{I}_k} (\langle e_I \sqcup e_I, \cdot \rangle - 2\langle \mathbb{E}_\mu[S_N(X)], \cdot \rangle + \|\mathbb{E}_\mu[S_N(X)]\|_2^2),$$

where $\mathcal{I}_k = \{1, \dots, d\}^k$ is the set of multi-indices of length k . It may also be rewritten in terms of the truncated signature kernel κ_N from (2.5) as

$$f(x) = \sqrt{\kappa_N(x, x) - 2\mathbb{E}_\mu[\kappa_N(X, x)] + \mathbb{E}_\mu[\kappa_N(X, X')]}, \quad (2.9)$$

where X' is an independent copy of $X \sim \mu$, and (2.9) can be evaluated using a kernel trick. After precomputing the terms that do not depend on x , each (kernelised and direct) evaluation of the score has the same cost as the ones discussed for one-class SVMs. One can generalise this construction by replacing the ambient ℓ_2 -norm with other norms. For example, a “variance norm” leads to a Mahalanobis-like distance in feature space. This perspective is developed in more detail in the next section. See also Cass et al. (2024) for a related kernelised treatment.

2.3 Conformance score

One of the most widely used nonparametric approaches to anomaly detection is the nearest neighbor (NN) algorithm (Bouman et al., 2024). Given a dataset of normal reference samples, the NN score of a new observation is defined by its distance to the closest reference point. For time series data, this requires specifying a metric on path space. A signature-based approach was introduced in Shao et al. (2020), where each path is embedded into a truncated signature, and distances are then computed in this feature space.

Formally, for a subset $C \subset \mathcal{X}$, the distance of a point $x \in \mathcal{X}$ to C is defined as

$$\text{dist}_{\mathcal{X}}(x, C) := \inf_{y \in C} \|x - y\|_{\mathcal{X}},$$

where $\|\cdot\|_{\mathcal{X}}$ is a metric on path space. Replacing paths by their signatures, distances are computed with respect to a chosen norm on the feature space \mathcal{H} . Given a measure μ on \mathcal{X} , choosing the *variance norm* (Definition A.1) associated to the measure $\nu = \mu \circ S^{-1}$ on \mathcal{H} , defined by

$$\|x\|_{\nu\text{-cov}} := \sup_{\substack{x^* \in \mathcal{H}^* \\ \text{Cov}_\nu(x^*, x^*) \leq 1}} x^*(x),$$

yields what Shao et al. (2020) refer to as the *conformance score*. In the case of a Gaussian reference distribution μ , the variance norm $\|\cdot\|_{\mu}$ coincides with the norm of the corresponding Cameron-Martin space \mathbb{H} . More details can be found in Section A.1 where we develop a general theory of conformance scores. Additionally, by the universal approximation property of signatures (Cass and Salvi, 2024, Theorem 1.4.7), the conformance score can itself be reformulated as a linear functional of the signature, thus fitting naturally within the framework introduced above.

3 Hypothesis testing for novelty detection on path space

Having introduced signature-based test statistics, we now turn to their use in hypothesis testing. To determine critical values, we derive analytical approximations to their quantiles, exploiting algebraic properties of signatures together with concentration inequalities to obtain non-asymptotic bounds.

3.1 Sample-free quantile estimates via expected signature

The baseline approach to quantile estimation is Monte Carlo sampling, where the α -quantile is obtained by inverting the empirical distribution function. The resulting estimator, given by the $\lceil \alpha n \rceil^{th}$ order statistic converges at a rate $\mathcal{O}(n^{-1/2})$ (Serfling, 2009). In risk management problems, it is well known that the quantile admits a variational characterisation as the minimiser of a conditional value-at-risk functional. In financial applications, CVaR is often preferred to the value-at-risk (VaR), the quantile itself, due to its convexity and variational properties. Building on this observation, we show that smooth polynomial CVaR surrogates of the pushforward of a measure μ on path space under the evaluation map $x \mapsto \langle w, S_N(x) \rangle$ admits a deterministic characterisation in terms of the expected signature.

Definition 3.1 (CVaR) *For any real-valued random variable Z with law \mathbb{P} , the conditional value-at-risk $\text{CVaR}_\alpha(Z)$ is defined as the average of the α -tail of the probability distribution of Z for a specified confidence level $\alpha \in [0, 1]$, that is*

$$\text{CVaR}_\alpha(Z) := \mathbb{E}[Z | Z \geq \text{VaR}_\alpha(Z)],$$

where $\text{VaR}_\alpha(Z) := \inf\{r | P_Z(z \leq r) \geq \alpha\}$ is the α -quantile of Z .

Its variational formulation (Rockafellar and Uryasev, 2002) makes it possible to calculate both VaR and CVaR simultaneously:

Lemma 3.2 *For any real-valued random variable Z and any $\alpha \in [0, 1)$, we have*

$$\text{CVaR}_\alpha(Z) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{\mathbb{E}[[Z - \eta]^+]}{1 - \alpha} \right\}.$$

Let $n \in \mathbb{N}$ and $Q_n : \mathbb{R} \rightarrow \mathbb{R}$ be a polynomial approximation of the max-function $[\cdot]^+$ on any compact interval $[-K, K] \subset \mathbb{R}$ (such an approximation is guaranteed by the Stone-Weierstrass theorem in the compact-open topology) and consider the functions $f_\alpha, f_\alpha^n : \mathbb{R} \rightarrow \mathbb{R}$ defined for all $\rho \in \mathbb{R}$ by

$$f_\alpha(\rho) := \rho + \frac{\mathbb{E}[[Z - \rho]^+]}{1 - \alpha} \quad \text{and} \quad f_\alpha^n(\rho) := \rho + \frac{\mathbb{E}[Q_n(Z - \rho)]}{1 - \alpha},$$

so that the CVaR and the smoothed CVaR read

$$\text{CVaR}_\alpha(Z) = \min_{\rho \in \mathbb{R}} f_\alpha(\rho) \quad \text{and} \quad \text{CVaR}_\alpha^n(Z) = \min_{\rho \in [-K, K]} f_\alpha^n(\rho).$$

By the shuffle product property (2.2) of the signature—the multiplication of two linear forms on the signature is still a linear form on the signature—the stochastic optimisation for $\text{CVaR}_\alpha^n(\langle w, S(X) \rangle)$ may be recast as the minimisation of a deterministic function of the expected signature of the process X .

Theorem 3.3 *Let $Q_n(x) = \sum_{i=0}^n a_i x^i$ be a polynomial of degree n and X a stochastic process with law μ and satisfying $\mathbb{E}_\mu[S(X)] < \infty$. For any linear functional $w \in (T^N(\mathbb{R}^d))^*$,*

$$\mathbb{E}_\mu[Q_n(\langle w, S(X) \rangle - \rho)] = \langle Q_n^{\sqcup}(w - \rho \mathbf{1}), \mathbb{E}_\mu[S(X)] \rangle,$$

where $Q_n^{\sqcup} : (T^N(\mathbb{R}^d))^* \rightarrow (T^{nN}(\mathbb{R}^d))^*$ is defined by $Q_n^{\sqcup}(\ell) := \sum_{i=0}^n a_i \ell^{\sqcup i}$ and

$$\text{CVaR}_\alpha^n(\langle w, S(X) \rangle) = \min_{\rho \in [-K, K]} \sum_{m=0}^n b_m \rho^m,$$

where the coefficients b_m are given explicitly by

$$b_m = \delta_{\{m=1\}} + \frac{1}{1-\alpha} \sum_{i=m}^n a_i \binom{i}{m} (-1)^m \left\langle w^{\sqcup(i-m)}, \mathbb{E}_\mu[S(X)] \right\rangle.$$

Proof First, we rewrite the CVaR objective using the shuffle product property (2.2) as

$$\begin{aligned} f_\alpha^n(\rho) &= \rho + \frac{\mathbb{E}_\mu[Q_n(\langle w, S(X) \rangle - \rho)]}{1-\alpha} \\ &= \rho + \frac{\mathbb{E}_\mu[Q_n(\langle w - \rho \mathbf{1}, S(X) \rangle)]}{1-\alpha} \\ &= \rho + \frac{\mathbb{E}_\mu[\sum_{i=0}^n a_i (\langle w - \rho \mathbf{1}, S(X) \rangle)^i]}{1-\alpha} \\ &= \rho + \frac{\mathbb{E}_\mu[\sum_{i=0}^n a_i \langle (w - \rho \mathbf{1})^{\sqcup i}, S(X) \rangle]}{1-\alpha} \\ &= \rho + \frac{\langle \sum_{i=0}^n a_i (w - \rho \mathbf{1})^{\sqcup i}, \mathbb{E}_\mu[S(X)] \rangle}{1-\alpha} \\ &= \rho + \frac{\sum_{i=0}^n a_i \sum_{k=0}^i \binom{i}{k} (-\rho)^{i-k} \langle w^{\sqcup k}, \mathbb{E}_\mu[S(X)] \rangle}{1-\alpha}. \end{aligned}$$

Then, f_α^n can be rewritten as a polynomial $\sum_{m=0}^n b_m \rho^m$ in ρ with coefficients b_m given by

$$b_m = \delta_{m=1} + \frac{1}{1-\alpha} \sum_{i=m}^n a_i \binom{i}{m} (-1)^m \left\langle w^{\sqcup(i-m)}, \mathbb{E}_\mu[S(X)] \right\rangle.$$

■

Tsyurmasto et al. (2014) showed that one-class SVMs can be interpreted as minimising the regularised CVaR (see Section A.2):

$$w^* = \arg \min_{w \in \mathcal{H}} \left\{ \text{CVaR}_\alpha(-\langle w, \varphi(X) \rangle_{\mathcal{H}}) + \frac{1}{2} \|w\|_{\mathcal{H}}^2 \right\}. \quad (3.1)$$

If φ is chosen as the truncated signature S_N and CVaR is replaced by a smooth polynomial surrogate, Theorem 3.3 allows us to reformulate the stochastic quadratic programming problem in (3.1) into a more tractable optimisation task, that amounts to minimising a polynomial in the variable ρ with coefficients expressed in terms of the expected signature $\mathbb{E}_\mu[S(X)]$ of the stochastic process X .

3.2 Probabilistic quantile estimates via transportation-cost inequalities

Throughout this section, we fix $\gamma \in (0, 1)$ and denote by $C^\gamma([0, 1]; \mathbb{R}^d)$ the Banach space of γ -Hölder continuous paths $x : [0, 1] \rightarrow \mathbb{R}^d$ endowed with the norm

$$\|x\|_\gamma := \sup_{t \in [0, 1]} |x(t)| + \sup_{t \neq s} \frac{|x(t) - x(s)|}{|t - s|^\gamma}.$$

For technical reasons, we define $C^\gamma([0, 1]; \mathbb{R}^d)$ as the completion of smooth paths under the Hölder norm which is a Polish space (separable, completely metrisable topological space). We will often write C^γ instead of $C^\gamma([0, 1]; \mathbb{R}^d)$ when there is no risk of confusion. The space of Borel probability measures on a Polish space \mathcal{X} is denoted by $\mathcal{P}(\mathcal{X})$. In order to derive error bounds for hypothesis testing, we will assume that reference measures μ , under the null hypothesis, satisfy Transportation-Cost Inequalities (TCIs), which we now recall, and we refer the reader to (Gasteratos and Jacquier, 2023) for more details on these functional inequalities.

Definition 3.4 *Let \mathcal{X} be a Polish space, $c : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty]$ a measurable function and $\mu, \nu \in \mathcal{P}(\mathcal{X})$.*

1. *The transportation cost between μ and ν with respect to the cost function c reads*

$$W_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathcal{X} \times \mathcal{X}} c(x, y) d\pi(x, y),$$

where $\Pi(\mu, \nu)$ is the collection of couplings between μ and ν :

$$\Pi(\mu, \nu) := \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : [\pi]_1 = \mu, [\pi]_2 = \nu \right\}.$$

2. *The relative entropy of ν with respect to μ is given by*

$$\mathbf{H}(\nu \mid \mu) := \begin{cases} \int_{\mathcal{X}} \log \left(\frac{d\nu}{d\mu} \right) d\nu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

The following TCIs generalise Talagrand's classical transportation-cost inequality (Talagrand (1996)). These functional inequalities imply heavier-than-Gaussian tail bounds and lead to informative error bounds for hypothesis tests that are expressed with respect to relative entropies between the null and alternative hypothesis measures.

Definition 3.5 *Let \mathcal{X} be a Polish space, $\mu \in \mathcal{P}(\mathcal{X})$, $c : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty]$ a measurable function with $c(x, x) = 0$ for all $x \in \mathcal{X}$ and $a : [0, \infty] \rightarrow [0, \infty]$ a lower semicontinuous function with $a(0) = 0$. We say that μ satisfies the (a, c) -TCI (and write $\mu \in \mathcal{T}_a(c)$) with cost function c and deviation function a if, for all $\mathcal{P}(\mathcal{X}) \ni \nu \ll \mu$,*

$$a\left(W_c(\mu, \nu)\right) \leq \mathbf{H}(\nu \mid \mu). \tag{3.2}$$

For the rest of this section we make the following standing assumption:

Assumption 3.6 For $\gamma \in (0, 1)$ we take $\mathcal{X} = C^\gamma([0, 1]; \mathbb{R}^d)$. The reference probability measure $\mu \in \mathcal{P}(C^\gamma([0, 1]; \mathbb{R}^d))$ satisfies the (a, c) -TCI with cost function

$$c(x, y) := \|x - y\|_\gamma^p, \quad \text{for all } x, y \in C^\gamma$$

and a continuous, strictly increasing deviation function $a : [0, \infty] \rightarrow [0, \infty]$ with $a(0) = 0$.

For a sample path X we consider the hypothesis test:

$$\begin{aligned} H_0 : X &\sim \mu, \\ H_1 : X &\sim \nu \neq \mu, \end{aligned} \tag{3.3}$$

where $\nu \in \mathcal{P}(C^\gamma([0, 1]; \mathbb{R}^d))$ is an arbitrary probability measure on \mathcal{X} . The test statistic is chosen to be the decision function $F : C^\gamma([0, 1]; \mathbb{R}^d) \rightarrow \{0, 1\}$ defined by

$$F(x) := \mathbb{1}_{\{\langle w, S_N(x) \rangle > r\}}(x), \quad \text{for } x \in C^\gamma([0, 1]; \mathbb{R}^d), N \in \mathbb{N},$$

where $w \in \mathcal{H}$ lies in the signature RKHS and solves the corresponding one-class SVM optimisation problem and $r \in \mathbb{R}$ is arbitrary but will later be chosen to be an α -quantile as in (A.1). This class of test statistics includes all of the examples described in Section A. The type-I and type-II errors are thus given by

$$P_I := \mathbb{P}\left(F(X) = 1 | X \sim \mu\right) = \mu\left(\langle w, S_N(x) \rangle > r\right), \tag{3.4}$$

$$P_{II} := \mathbb{P}\left(F(X) = 0 | X \sim \nu\right) = \nu\left(\langle w, S_N(x) \rangle \leq r\right). \tag{3.5}$$

For the rest of this section we assume that $r > 0$ and derive upper and lower bounds for the type-I and type-II errors in Theorems 3.9, 3.11 respectively. The case $r < 0$ will be discussed in Remark 3.13 below. Given w, r , Theorem 3.11 provides the upper bound for the type-I error:

$$P_I \leq C_2 \left[\exp \left\{ -\frac{C_1^2}{2} \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{2p} \vee \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{\frac{2p}{N}} \right\} \right],$$

where d is the dimension of the paths, C_1, C_2 are absolute constants that depend on the exponential moments of μ and the deviation function a is assumed to be super-quadratic near the origin (an assumption that is satisfied in all the examples of interest).

Example 1 (RDEs with Gaussian drivers) An important class of measures that satisfy (a, c) -TCIs is given by laws of solutions to Rough Differential Equations (RDEs) driven by Gaussian rough paths. In particular, let $\beta \in (\frac{1}{3}, 1)$ and X be a Gaussian process with almost surely β -Hölder continuous paths that has a natural geometric β -rough path lift \mathbf{X} .

If the corresponding Cameron-Martin space satisfies $\mathcal{H} \hookrightarrow C^{q\text{-var}}([0, T]; \mathbb{R}^d)$ for some $q > 0$, with $\frac{1}{q} + \beta > 1$, then the law $\mu \in \mathcal{P}(C^{1/\beta\text{-var}}([0, T]; \mathbb{R}^d))$ of the solution Y of the RDE

$$dY_t = V(Y_t)d\mathbf{X}_t,$$

where V is a Lipschitz vector field with constant $\gamma > 1/\beta$, satisfies a $\mathcal{T}_a(c)$ -inequality (Gasteratos and Jacquier, 2023, Theorem 3.3) with

$$c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{\frac{1}{\beta}\text{-var}}^{\frac{1}{q}} \quad \text{and} \quad a(t) = t^2 \wedge t^{2q}.$$

In view of (Gasteratos and Jacquier, 2023, Corollary 3.5), the latter implies the tail estimates in (Cass et al., 2013). While the results in this section do not depend on the particular choice of path topologies, the aforementioned TCI for solution laws of RDEs is valid on finite p -variation topologies as it leverages complementary Young regularity. Thus, all our results can be applied to this measure *mutatis mutandis* i.e. by substituting the Hölder space and norm in Assumption 3.6 and proofs by the corresponding $1/\beta$ -variation topology.

From the fact that $\mu \in \mathcal{T}_a(c)$ we will establish lower bounds for the type-II error. To do so, we rely on the following dual characterisation of the transportation cost W_c .

Lemma 3.7 *Let $p \in (0, 1]$, $\gamma \in (0, 1)$ as in Assumption 3.6 and*

$$\mathcal{C}^{p,\gamma} := \left\{ f : C^\gamma([0, 1]; \mathbb{R}^d) \rightarrow \mathbb{R} : \inf_{x \neq y \in C^\gamma} \frac{|f(x) - f(y)|}{\|x - y\|_\gamma^p} \leq 1 \right\}.$$

Then, for any $\mu, \nu \in \mathcal{P}(C^\gamma([0, 1]; \mathbb{R}^d))$ with finite first moments,

$$W_c(\mu, \nu) = \sup_{f \in \mathcal{C}^{p,\gamma}} \left\{ \mathbb{E}_\mu[f] - \mathbb{E}_\nu[f] \right\}.$$

Proof By Kantorovich duality (Gozlan and Léonard, 2010, Theorem 2.2),

$$W_c(\mu, \nu) = \sup \left\{ \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] \mid f \in L^1(\mu), g \in L^1(\nu), f(x) + g(y) \leq c(x, y) = \|x - y\|_\gamma^p \right\}.$$

For any fixed $g \in L^1(\nu)$, using triangle and Young's product inequalities, the function

$$f^*(y) := \inf_{x \in C^\gamma} \left\{ \|x - y\|_\gamma^p - g(x) \right\}$$

satisfies

1. With $(1/p)^* := (1 - p)^{-1}$ and for all $y \in C^\gamma([0, 1]; \mathbb{R}^d)$ we have $f^*(y) \leq \|y\|_\gamma^p - g(0) \leq p\|y\|_\gamma + \frac{1}{(1/p)^*} + |g(0)|$, so that $f^* \in L^1(\mu)$ since μ has a finite first moment;
2. For any $x_2 \neq x_1 \in C^\gamma$, $|f^*(x_2) - f^*(x_1)| \leq \|x_2 - x_1\|_\gamma^p$, so that $f^* \in \mathcal{C}^{p,\gamma}$. Indeed, for any $\delta > 0$ there exists $\bar{x} \in C^\gamma$ such that $f^*(x_1) > \|\bar{x} - x_1\|_\gamma^p - g(\bar{x}) - \delta$. Thus,

$$f(x_2) - f(x_1) \leq \|\bar{x} - x_2\|_\gamma^p - g(\bar{x}) - \|\bar{x} - x_1\|_\gamma^p + g(\bar{x}) + \delta \leq \|x_1 - x_2\|_\gamma^p + \delta,$$

since the inequality $x^p - y^p \leq |x - y|^p$ holds for real numbers $x, y \geq 0$; we also used the triangle inequality for the norm $\|\cdot\|_\gamma$ and the monotonicity of the function $x \mapsto x^p$. Interchanging the roles of x_1, x_2 we obtain the reverse inequality

$$f(x_2) - f(x_1) > -\|x_2 - x_1\|_\gamma^p;$$

3. Out of all functions $f \in \mathcal{C}^{p,\gamma}$ that satisfy $f(x) + g(y) \leq c(x, y)$ we have $\mathbb{E}_\mu[f] \leq \mathbb{E}_\mu[f^*]$.

From these facts we deduce that

$$\begin{aligned} W_c(\mu, \nu) &= \sup_{g \in \mathcal{C}^{p,\gamma}} \sup_{\substack{f \in \mathcal{C}^{p,\gamma} \\ f \oplus g \leq c}} \left\{ \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] \right\} \\ &= \sup_{g, f \in \mathcal{C}^{p,\gamma}} \sup_{f \oplus g \leq c} \left\{ \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] \right\} \\ &= \sup_{f \in \mathcal{C}^{p,\gamma}} \sup_{\substack{g \in \mathcal{C}^{p,\gamma} \\ f \oplus g \leq c}} \left\{ \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] \right\}. \end{aligned} \tag{3.6}$$

Now, for any fixed $f \in \mathcal{C}^{p,\gamma}$, we can optimise the last equality by taking

$$g(y) = g^*(y) := \inf_{x \in C^\gamma} \{ \|x - y\|_\gamma^p - f(x) \}$$

as before. On the one hand, $g^*(y) \leq -f(y)$ by definition of g^* ; on the other hand, since $f \in \mathcal{C}^{p,\gamma}$ we must have, for all x, y , $\|x - y\|_\gamma^p - f(x) + f(y) \geq 0$. This implies $g^*(y) = \inf_{x \in C^\gamma} \{ \|x - y\|_\gamma^p - f(x) \} \geq -f(y)$ namely $g^* = -f$. Returning to (3.6) we obtain

$$W_c(\mu, \nu) = \sup_{f \in \mathcal{C}^{p,\gamma}, g = g^*} \left\{ \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] \right\} = \sup_{f \in \mathcal{C}^{p,\gamma}} \left\{ \mathbb{E}_\mu[f] - \mathbb{E}_\nu[f] \right\},$$

which completes the proof. ■

With this characterisation of W_c we derive a lower bound for the type-II error using the following:

Corollary 3.8 *Let μ satisfy Assumption B.2. Then for any other measure $\nu \in \mathcal{P}(C^\gamma)$, with finite first moments and any $f \in \mathcal{C}^{p,\gamma}$,*

$$\mathbb{E}_\nu[f] \leq a^{-1} \left(\mathbf{H}(\nu|\mu) \right) + \mathbb{E}_\mu[f].$$

Proof The bound follows directly by the symmetry of W_c , Lemma 3.7 and the fact that $\mu \in \mathcal{T}_a(c)$ if and only if $a(W_c(\nu, \mu)) \leq \mathbf{H}(\nu|\mu)$ per Definition 3.5 ■

We are now ready to state novel type-II errors for the hypothesis test (3.3) in terms of the relative entropy.

Theorem 3.9 (Type-II error) *With p, γ, c, μ as in Assumption 3.6 and $w, N \in \mathbb{N}, r > 0$ as in (3.5) and for any $\nu \in \mathcal{P}(C^\gamma)$ with finite first moment, the following bounds hold:*

$$P_{\text{II}} \geq 1 - \left(\frac{\|w\|_{\mathcal{H}}}{r} \right)^{\frac{p}{N}} C d^p N^{\frac{1}{2Np}} \left\{ 1 - \frac{1}{N} + \frac{1}{N} \left(a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \right) \right\},$$

if $a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \leq 1$, and

$$P_{\text{II}} \geq 1 - \left(\frac{\|w\|_{\mathcal{H}}}{r} \right)^{\frac{p}{N}} C d^p N^{\frac{1}{2Np}} \left\{ a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \right\}, \text{ if } a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] > 1,$$

where d is the dimension of the path X and $C > 0$ a constant independent of d .

Proof By Chebyshev's inequality

$$P_{\text{II}} = 1 - \nu \left(\langle w, S_N(x) \rangle > r \right) \geq 1 - \left(\frac{\|w\|_{\mathcal{H}}}{r} \right)^{\frac{p}{N}} \mathbb{E}_\nu \left[|S_N(X)|^{\frac{p}{N}} \right]. \quad (3.7)$$

In view of (Boedihardjo, 2018, Section 1) and references therein, we can obtain a pathwise estimate for the signature, truncated at level N , of any d -dimensional, γ -Hölder path X with $\gamma \in (0, \frac{1}{2}]$. The latter takes the form:

$$\begin{aligned} |S_N(X)|^2 &\leq \sum_{k=1}^N \left(\sum_{|w|=k} S_{0,1}^w(X) \right)^2 \\ &\leq \sum_{k=1}^N C^2(k) \|X\|_\gamma^{2k} \left(\sum_{|w|=k} \frac{1}{\Gamma(k\gamma + 1)} \right)^2 \leq \|X\|_\gamma^2 \vee \|X\|_\gamma^{2N} \sum_{k=1}^N C^2(k) d^{2k}, \end{aligned}$$

where $|w|$ is the length of a word w , Γ denotes the Gamma function and $C(k) = \mathcal{O}((1+c)^k)$ is an increasing polynomially growing function in k . Hence the last estimate yields

$$|S_N(X)|^2 \leq \|X\|_\gamma^2 \vee \|X\|_\gamma^{2N} N d^{2N} (1 + \rho)^N, \quad (3.8)$$

for some constant $\rho > 0$. Taking expectations it follows that

$$\begin{aligned} \mathbb{E}_\nu \left[|S_N(X)|^{\frac{p}{N}} \right] &\leq d^p (1 + \rho)^p N^{\frac{p}{2N}} \mathbb{E}_\nu \left[\|X\|_\gamma^{\frac{p}{N}} \vee \|X\|_\gamma^p \right] \\ &\leq C d^p N^{\frac{p}{2N}} \mathbb{E}_\nu \left[\|X\|_\gamma^{\frac{p}{N}} \right] \vee \mathbb{E}_\nu \left[\|X\|_\gamma^p \right], \end{aligned}$$

for some unimportant constant $C > 0$. By Young's product inequality, we have

$$\mathbb{E}_\nu \left[\|X\|_\gamma^{\frac{p}{N}} \right] \leq \frac{1}{\frac{N}{N-1}} + \frac{1}{N} \mathbb{E}[\|X\|_\gamma^p] = 1 - \frac{1}{N} + \frac{1}{N} \mathbb{E}_\nu[\|X\|_\gamma^p].$$

Now, notice that the function $f(x) = \|x\|_\gamma^p$ is in $\mathcal{C}^{p,\gamma}$. In view of Corollary 3.8 then

$$\mathbb{E}_\nu \left[\|X\|_\gamma^{\frac{p}{N}} \right] \leq 1 - \frac{1}{N} + \frac{1}{N} \left(a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \right).$$

Combining all these estimates we see that $\mathbb{E}_\nu \left[|S_N(X)|^{\frac{p}{N}} \right]$ is bounded from above by

$$C d^p N^{\frac{p}{2N}} \max \left\{ 1 - \frac{1}{N} + \frac{1}{N} \left(a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \right), \left(a^{-1}(\mathbf{H}(\nu|\mu)) + \mathbb{E}_\mu[\|X\|_\gamma^p] \right) \right\}.$$

The proof then follows by combining this with (3.7). ■

Remark 3.10 *A few observations on the lower bound for the type-II error in Theorem 3.9:*

1. *If the measure ν from the alternative hypothesis is not absolutely continuous with respect to μ then $\mathbf{H}(\nu|\mu) = \infty$ and, since a^{-1} is (typically) increasing to ∞ , the lower bound is then uninformative. Nevertheless, we believe that the regime in which $\mathbf{H}(\nu|\mu) < \infty$ is more interesting for application purposes. Indeed, if ν, μ are absolutely continuous laws on path space then, a priori, it is more challenging to tell them apart (or detect novel observations in terms of these two alternatives). It is precisely this setting in which the type-II lower bound becomes informative. Nevertheless, our numerical experiments work well even in settings where the two alternatives are mutually singular; see Section 4.1 for more details.*
2. *In estimating the type-II error, we made a choice to raise the signature to p/N (below (3.8)). This choice was made to avoid exponential growth in the dimension d which appears on the right-hand side of (3.8).*

We turn to the analysis of type-I errors (3.4). Assuming for now that μ is a Gaussian measure, it is possible to obtain estimates on the probability of the rejection region (1.1) under H_0 . In particular, for $\mathcal{X} = C^{p\text{-var}}([0, T]; \mathbb{R}^d)$, $\mu \in \mathcal{P}(\mathcal{X})$ a Gaussian probability measure on path space and for some $N \in \mathbb{N}$ consider the truncated signature feature map $\varphi = S_N$ of level N . The probability that a new sample point from μ falls in the rejection region satisfies

$$\mu(\Omega_r) = \mu\left[\langle w, S_N(X) \rangle > r\right] \leq \exp\left\{-C\left(\frac{r}{\|w\|_{\mathcal{H}}}\right)^{\frac{2}{N}}\right\}.$$

In turn, for a given significance level $\alpha \in (0, 1)$, the following estimate on r holds:

$$r \geq C^{-1}\|w\|_{\mathcal{H}} \log(1/\alpha)^{\frac{N}{2}}.$$

Such estimates can be performed not just for Gaussian measures μ but also for laws of Gaussian functionals such as solutions of RDEs or in general measures that satisfy transportation-cost inequalities. A broad list of examples of such measures can be found in (Gasteratos and Jacquier, 2023); see also Example 1 above.

Our next result provides type-I error upper bounds and consequently estimates on the threshold r of the rejection region Ω_r (1.1) at fixed significance levels.

Theorem 3.11 (Type-I error) *Let $\alpha \in (0, 1)$, $N \in \mathbb{N}$, $\gamma \in (0, 1)$, $p \in (0, 1]$ and $S_N : C^\gamma([0, 1]; \mathbb{R}) \rightarrow T^{(N)}(\mathbb{R})$ the truncated signature of level N . Let $\mu \in \mathcal{P}(C^\gamma([0, 1]; \mathbb{R}))$ be a probability measure satisfying Assumption 3.6. If there exist $t_0, C > 0$ such that for all $t < t_0$ $a(t) \geq Ct^2$, then the following holds:*

For all $w \in T^{(N)}(\mathbb{R})$ and with probability $\alpha \in (0, 1)$ over the draw of a random sample from μ , a sample path X that satisfies

$$\langle w, S_N(X) \rangle > r^* := \|w\|_{\mathcal{H}} \sqrt{Nd^N} C(N) \max\left\{\left[\frac{2}{C_1} \log\left(\frac{C_2}{\alpha}\right)\right]^{\frac{1}{2p}}, \left[\frac{2}{C_1} \log\left(\frac{C_2}{\alpha}\right)\right]^{\frac{N}{2p}}\right\} \quad (3.9)$$

falls in the rejection region Ω_{r^*} (1.1). The constants C_1, C_2 do not depend on N and satisfy

$$C_1 \leq \left(2(\mathbb{E}_\mu \|X\|_\gamma^p + C)\right)^{-\frac{1}{2}} \quad \text{and} \quad C_2 = \mathbb{E}_\mu \left[\exp \left(\frac{C_1^2}{2} \|X\|_\gamma^{2p} \right) \right].$$

Proof The pathwise estimate (3.8) furnishes

$$\begin{aligned} \alpha = \mu(\Omega_r) &\leq \mu \left(|S_N(X)| \|w\|_{\mathcal{H}} > r \right) \leq \mu \left(\|X\|_\gamma \vee \|X\|_\gamma^N > \left(\frac{r}{\sqrt{N} d^N (1+\rho)^{N/2} \|w\|_{\mathcal{H}}} \right) \right) \\ &\leq \mu \left(\|X\|_\gamma > \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right) \vee \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{\frac{1}{N}} \right). \end{aligned}$$

From the assumptions on μ, a , (Gasteratos and Jacquier, 2023, Corollary 2.6(ii)) and Chebyshev's inequality, the latter is bounded from above by

$$C_2 \left[\exp \left\{ -\frac{C_1^2}{2} \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{2p} \vee \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{\frac{2p}{N}} \right\} \right],$$

from which we deduce that

$$\log(\alpha/C_2) \leq -\frac{C_1^2}{2} \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{2p} \vee \left(\frac{r}{\sqrt{N} d^N C^{N/2} \|w\|_{\mathcal{H}}} \right)^{\frac{2p}{N}}.$$

Solving this inequality for r , we are led to the desired conclusion. ■

Remark 3.12 (Tradeoffs related to N) From the estimate (3.9) we observe that, at fixed significance level α , the rejection threshold $r^* = r^*(N)$ grows as the signature truncation level N grows. In particular, rejecting H_0 becomes harder since larger observations become more typical (indeed, tails of the signature become "fatter" as N increases). On the other hand, larger values of N lead to higher robustness for test statistics. Indeed, since linear maps of the signature are universal, larger values of N lead to larger classes of admissible test statistics on path space.

Remark 3.13 (Error bounds when $r < 0$) We assumed above $r > 0$ and derived upper/lower bounds for type-I/II errors respectively. In the case $r < 0$, it is easy to check that our techniques provide symmetric lower/upper bounds for type-I/II errors respectively.

4 Experiments

In this section, we evaluate the type-I error and the statistical power of signature-based test statistics, using synthetic anomalous diffusion data and real-world molecular biology data.

4.1 Anomalous diffusions

We consider the problem of distinguishing Brownian motion (BM) from BM perturbed by an impulsive spike occurring at a uniformly random time. This example, connected to the theory of Brownian fast points (Davis and Perkins, 1985), has been introduced by Davis and Monroe (1984) and Hitchcock (1992), and has also been empirically examined by Cass et al. (2024). For a given sample path X , we consider the hypothesis test

$$H_0 : X \sim \mu \text{ against } H_1 : X \sim \nu_\varepsilon,$$

with μ the law of Brownian motion on $[0, 2]$ and ν_ε that of a process with a spike. Following the presentation in (Hitchcock, 1992, Chapter 6), define the alternative process

$$X_t^\varepsilon = B_t + \varepsilon \sqrt{[t - \theta]^+} \wedge 1, \quad \text{for } t \in [0, 2],$$

where θ is a random time drawn from the uniform distribution on $[0, 1]$, $\varepsilon > 0$ controls the magnitude of the spike, and $[x]^+ = \max\{x, 0\}$. Our numerical studies show how our detection procedure—based on signature transforms of the paths—can distinguish between μ and ν_ε for different intensities ε using the Area Under the Receiver Operating Characteristic curve (AUROC) as performance metric. We expect that as ε increases, the spike gets more pronounced and produces a clearer non-Brownian signature, leading to higher detection power. When $\varepsilon \approx 0$, the spike is subtle so detection is more challenging.

Hitchcock (1992) established that there is a critical value for the intensity parameter ε . When $\varepsilon < \sqrt{8}$, the measure corresponding to the process with a spike remains absolutely continuous with respect to the measure of standard Brownian motion. Conversely, when $\varepsilon > \sqrt{8}$ the two measures become singular. In our experiments, using the distance to the expected signature as test statistic, we do not observe an abrupt phase transition precisely at $\varepsilon = \sqrt{8}$. Instead, the AUROC improves gradually as ε increases (Figure 1a).

Next, to evaluate the type-I error rate and the power, we simulate $J = 100$ researchers. Each researcher j has an independent fixed dataset D_j of $n = 2000$ Brownian motion sample paths and $L = 50$ test sets $D_{j,l}^{\text{test}}$ (for $l = 1, \dots, L$) of $n_{\text{test}} = 1000$ Brownian motion sample paths, with 10% of outliers (Brownian motion perturbed by a spike). Each researcher divides its reference dataset D_j into 1000 sample paths to estimate the expected signature of Brownian motion and 1000 sample paths to estimate the p-values of all observations across all test datasets $D_{j,l}$. To measure the false discovery rate (FDR), the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) is applied to all p-values in $D_{j,l}$ at level $\alpha = 0.10$, the false positive rate (FPR) is then estimated, and averaged over the L test datasets. Similarly, a statistical estimate for the power is obtained. To estimate the false positive rate, the raw p-values are thresholded at level $\alpha = 0.01$.

As can be seen on Figures 1b and 1d, which show the FDR and FPR for different values of the signal strength ε^2 , both methods control the marginal FDR and FPR at the desired levels. We also compared to a parametric approach (Weibull), where we fit a curve $A \exp(-Br^{2/N})$ to the tail of the reference measure estimated with a relatively large number of samples, leveraging the results from Section 3.2. As can be seen on Figure 1b, this parametric approach controls the conditional FDR and FPR for a greater proportion of researchers. As expected, the false positive rate exhibits a lower variance (Figure 1d) when the Weibull parametric approach is used.

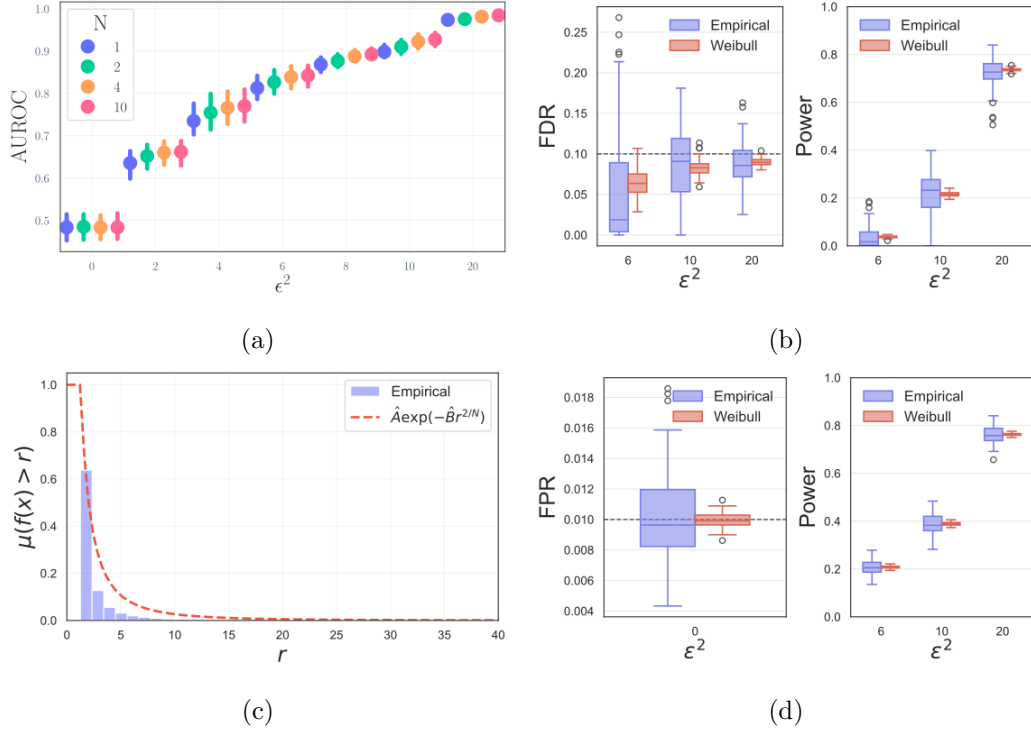


Figure 1: **Brownian motion perturbed by a spike.** (a) AUROC as a function of the spike intensity for the distance to the expected signature. (b) Multiple testing false discovery rate and power at level $\alpha = 0.1$ comparing empirical p-values (from 1,000 samples) with p-values obtained from a Weibull tail-bound fitted using 100,000 samples. (c) Weibull tail-bound fit (d) Single hypothesis false positive rate and power at level $\alpha = 0.01$.

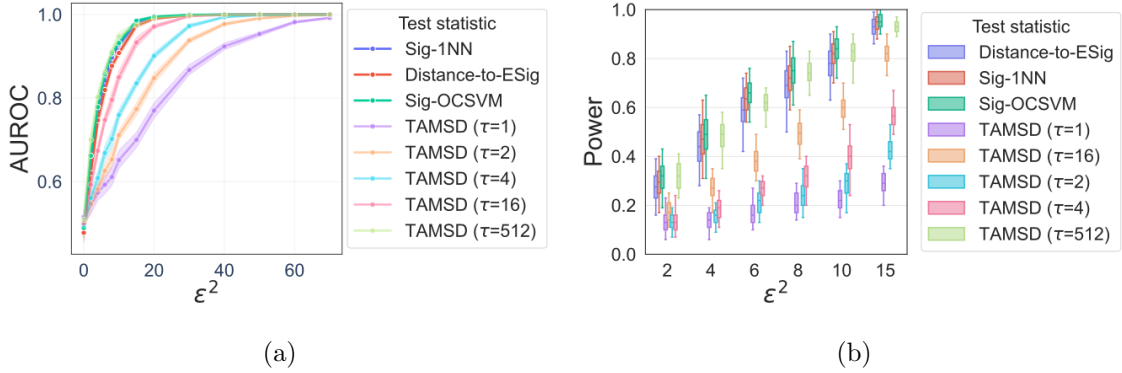


Figure 2: **Comparison of different test statistics.** Comparison of one-class SVM, conformance score, distance to the expected signature, and TAMSD with $\tau \in \{1, 2, 4, 16, 512\}$.

Finally, we compare our signature-based test statistics with another test statistic for detecting anomalous diffusions from the literature (Sikora et al., 2017) given by the time-

averaged mean square displacement (TAMSD)

$$M_N(\tau) := \frac{1}{N-\tau} \sum_{j=1}^{N-\tau} (X(j+\tau) - X(j))^2. \quad (4.1)$$

The TAMSD scales as $M_N(\tau) \sim \tau^\alpha$ with $\alpha = 2H$ for fractional Brownian motion with Hurst index $H \in (0, 1)$. Brownian motion has a linear power-law growth of the mean squared displacement in the course of time.

4.2 Anomaly detection in RNA direct sequencing data

In molecular biology, a central challenge is to detect non-canonical nucleotides on RNA molecules, i.e. bases chemically modified relative to the four standard nucleotides (A, U, G, C). Mapping such modifications has become a major research focus, as they have been found to modulate crucial cellular processes and are increasingly linked to human diseases, ranging from neurodevelopmental disorders to cancer. Recently, significant progress has been achieved by leveraging Nanopore direct sequencing data (White and Hesselberth, 2022; Furlan et al., 2021). Nanopore sequencing transduces long RNA polymers into an electrical signal and chemical modifications have been shown to induce characteristic alterations in the measured signal. Most prior approaches have focused on training neural network based classifiers on known modification types. However, since many modifications remain unknown or poorly characterised, this is also naturally cast as a novelty detection task Lemerrier et al. (2025).

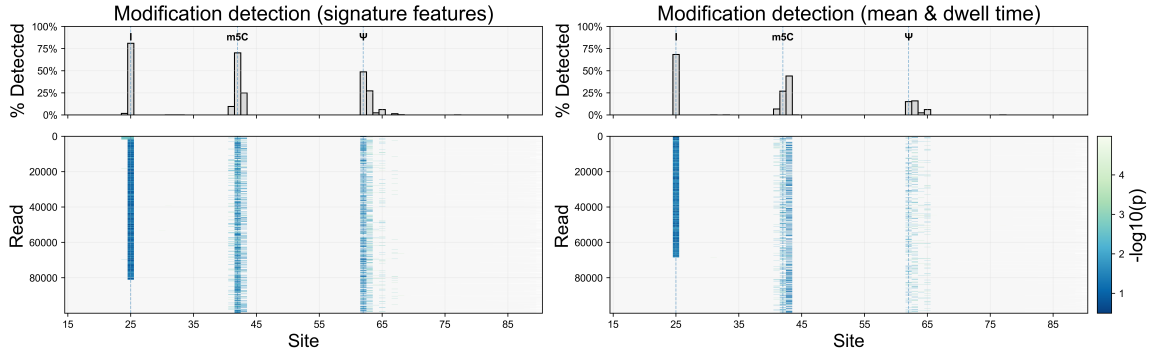


Figure 3: **Modification detection in nanopore reads with one-class SVM.** Bottom: per-read p-values (with multiple testing correction) at each site; non-significant p-values at level 0.20 are rendered light grey. Top: for each site, the proportion of significant reads. *Left*: signature features. *Right*: mean-current and dwell-time features.

Here, we analyse short synthetic oligonucleotides, i.e., RNA molecules chemically synthesised in a laboratory. In this controlled setting, three distinct chemical modifications were introduced at defined positions within a 100-nucleotide sequence (Leger et al., 2021)

AUACUCGACAUAGAUAGGACUCUUU(I) GCUAGUGAACCCUAGCCUC(m5C) GGAGACAGGUCGCGACCUGUG(Ψ) AGAUGAGAGAACUGAGUGCACAAAAAAAAAAAAA

The raw nanopore direct sequencing data were obtained from the European Nucleotide Archive (accession number PRJEB44511). It contains an unmodified sample and a sample where the three modifications have been deposited. Basecalling and signal processing were performed using Dorado¹, read alignment was carried out with minimap2 (Li, 2018), and event-level signal segmentation was obtained with Uncalled4 (Kovaka et al., 2025), yielding segmented time series representations for our analysis.

We compared a one-class SVM using signatures truncated at level $N = 6$ after applying two path transforms (the *time augmentation* and the *invisibility-reset* transforms (Lyons and McLeod, 2025) and a one-class SVM trained on two standard nanopore features given by the signal duration (commonly referred to as the *dwell time*) and signal mean value. We fitted the one-class SVM using 3 000 reads from the unmodified sample for each site, and computed the empirical p-values using 100 000 other unmodified reads. Figure 3 shows that, at a Benjamini–Hochberg FDR level of 0.20 (with Storey correction), the signature-feature model yields consistently higher recall for all three modification types, namely inosine (I), 5-methylcytosine (m5C), and pseudouridine (Ψ).

5 Conclusion

In this work, we have framed novelty detection on path space as a hypothesis testing problem based on signature statistics. By leveraging the transportation-cost inequalities of Gasteratos and Jacquier (2023), we established tail bounds for false positive rates that go beyond Gaussian measures to encompass laws of RDE solutions with smooth bounded vector fields, thereby enabling rigorous quantile and p-value estimates. Through the shuffle product structure, we obtained exact formulas for smooth surrogates of conditional value-at-risk (CVaR) in terms of expected signatures, which in turn motivated new one-class SVM algorithms optimising smooth CVaR objectives. We further derived lower bounds on type-II error under alternatives with finite first moment, leading to general power bounds in the absolutely continuous setting. Finally, our numerical experiments on synthetic anomalous diffusion data and real molecular biology data demonstrated both the validity of type-I error control and the practical effectiveness of signature-based testing procedures.

These results highlight the dual theoretical and algorithmic contributions of signature methods to statistical testing and novelty detection on path space. Future directions include extending the analysis to heavier-tailed drivers, developing computationally efficient estimators of expected signatures in high dimensions, and exploring applications to other domains where path-dependent anomalies play a central role.

Acknowledgments and Disclosure of Funding

Acknowledgements. This work was supported in part by EPSRC (NSFC) under Grant EP/S026347/1, in part by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

1. <https://github.com/nanoporetech/dorado>

Appendix A. Signature test statistics

We provide more details on several commonly-used anomaly scores on path space, which are given (or can be approximated) by an element of a signature kernel RKHS \mathcal{H} . These include distance-based score functions as well as solutions of one-class SVM optimisation problems. Throughout this section, \mathbb{H} will denote Hilbert spaces induced by the covariance of a measure μ . When μ is Gaussian, \mathbb{H} will be the Cameron-Martin space of μ . This choice of notation is made to distinguish the latter from the signature kernel RKHS.

A.1 Conformance score

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a normed vector space and $\mu \in \mathcal{P}(\mathcal{X})$ a Borel probability measure on \mathcal{X} . Based on Cochrane et al. (2020), we define two notions of conformance of a point x to a corpus $C \subset \mathcal{X}$: one based only on the topology of \mathcal{X} and one that leverages the structure of the measure μ .

Definition A.1

1. The **covariance** of μ induces a bilinear form on the topological dual \mathcal{X}^* as follows: For all $x^*, y^* \in \mathcal{X}^*$,

$$\text{Cov}_{\mu}(x^*, y^*) := \int_{\mathcal{X}} x^*(x) y^*(x) d\mu(x) = \mathbb{E}_{\mu}[x^* y^*].$$

2. The **variance norm** of $x \in \mathcal{X}$ is given by

$$\|x\|_{\mu} := \sup_{\text{Cov}_{\mu}(x^*, x^*) \leq 1} x^*(x).$$

3. For a Borel measurable set $C \in \mathcal{B}(\mathcal{X})$ and a point $x \in \mathcal{X}$, the **μ -conformance** of x to C is given by

$$\text{dist}_{\mu}(x, C) := \inf_{y \in C} \|x - y\|_{\mu},$$

where $\|\cdot\|_{\mu}$ denotes the variance norm of μ on \mathcal{X} .

4. For a Borel measurable set $C \in \mathcal{B}(\mathcal{X})$ and a point $x \in \mathcal{X}$, the **topological conformance** of x to C is given by

$$\text{dist}_{\mathcal{X}}(x, C) := \inf_{y \in C} \|x - y\|_{\mathcal{X}}.$$

Remark A.2 The topological conformance of x to C is nothing but the classical distance of a point x to a subset C of a metric space.

The case where μ is a *Gaussian* measure and \mathcal{X} a separable Banach space is special. One can then define a linear space \mathbb{H} contained in \mathcal{X} by

$$\mathbb{H} := \{x \in \mathcal{X} : \|x\|_{\mu} < \infty\}.$$

For every element $h \in \mathbb{H}$ there exists a representative $h^* \in \mathcal{X}^*$ such that $\|h\|_\mu = \text{Cov}_\mu(h^*, h^*)$ and moreover \mathbb{H} is a Hilbert space with inner product given by

$$\langle h_1, h_2 \rangle_{\mathbb{H}} := \text{Cov}_\mu(h_1^*, h_2^*).$$

The latter in fact shows that $\|\cdot\|_{\mathbb{H}}$ coincides with the variance norm $\|\cdot\|_\mu$ above. The proof of the following result is contained in (Hairer, 2009, Section 4.2), in particular Exercise 4.38 therein.

Proposition A.3 *Let \mathcal{X} be a separable Banach space and $\mu \in \mathcal{P}(\mathcal{X})$ a Gaussian measure. Then the variance norm $\|\cdot\|_\mu$ is equal to the norm $\|\cdot\|_{\mathbb{H}}$ of the Cameron-Martin Hilbert space $\mathbb{H} \subset \mathcal{X}$ associated to μ .*

The following inequality explains why the conformance defined above is appropriate for detecting anomalies in data that is assumed to come from a Gaussian distribution.

Theorem A.4 (TSB inequality) *Let μ be a Gaussian measure on a separable Banach space \mathcal{X} . Let $C \in \mathcal{B}(\mathcal{X})$, $\hat{c} := \Phi^{-1}(\mu(C))$ and $\bar{\Phi} = 1 - \Phi$, where Φ is the distribution function of a standard Normal distribution. Then for all $r \geq 0$,*

$$\mu\left(x \in \mathcal{X} : \text{dist}_\mu(x, C) \geq r\right) \leq \bar{\Phi}(r + \hat{c}).$$

Moreover, for $r \geq -\hat{c}$ we have

$$\mu\left(x \in \mathcal{X} : \text{dist}_\mu(x, C) \geq r\right) \leq \exp\left\{-\frac{(r + \hat{c})^2}{2}\right\}.$$

Proof This is a straightforward reformulation of the TSB inequality in (Friz and Hairer, 2020, Theorem 11.6). In particular, by the triangle inequality, along with the fact that the distance is nonnegative, we have

$$\begin{aligned} \mu\left(x \in \mathcal{X} : \text{dist}_\mu(x, C) \geq r\right) &\leq \mu\left(x \in \mathcal{X} : \text{dist}_\mu(x, C) \geq r\right) \\ &\leq \mu\left(x \in \mathcal{X} : \forall y \in C, \|x - y\|_\mu = \|x - y\|_{\mathbb{H}} \geq r\right) \\ &\leq \mu\left(x \in \mathcal{X} : x \notin C + rB_{\mathbb{H}}\right) \leq \bar{\Phi}(\hat{c} + r), \end{aligned}$$

where $B_{\mathbb{H}}$ denotes the unit ball in \mathbb{H} and the last inequality follows from the TSB inequality mentioned above. The second assertion then follows by the elementary bound $\bar{\Phi}(z) \leq e^{-z^2/2}$ which holds for all $z \geq 0$. ■

Remark A.5 *As mentioned in (Cochrane et al., 2020), if the corpus has probability at least $\frac{1}{2}$ ($\mu(C) \geq \frac{1}{2}$), then $\hat{c} \geq \Phi^{-1}(\frac{1}{2}) = 0$. Therefore, the probability that the conformance of a (random) sample point X to C exceeds a threshold r is bounded above by $e^{-r^2/2}$. In fact, in this case, the second moment of the μ -conformance can be bounded uniformly over \hat{c} :*

$$\begin{aligned}\mathbb{E}_\mu [\text{dist}_\mu^2(X, C)] &= 2 \int_0^\infty r \cdot \mu(x : \text{dist}_\mu(x, C) > r) dr \\ &\leq 2 \int_0^\infty r \bar{\Phi}(\hat{c} + r) dr \leq 2 \int_0^\infty r \bar{\Phi}(r) dr = \frac{1}{2}.\end{aligned}$$

On the other hand, if $\mu(C) < \frac{1}{2}$, then $\hat{c} < 0$. In this case, the TSB inequality provides a \hat{c} -dependent upper bound for the second moment of the μ -conformance:

$$\begin{aligned}\mathbb{E}_\mu [\text{dist}_\mu^2(X, C)] &= 2 \int_0^\infty r \cdot \mu(x : \text{dist}_\mu(x, C) > r) dr \\ &\leq 2 \int_0^{-\hat{c}} r \bar{\Phi}(\hat{c} + r) dr + 2 \int_{-\hat{c}}^\infty r \bar{\Phi}(\hat{c} + r) dr \\ &\leq 2 \int_0^{-\hat{c}} r dr + 2 \int_0^\infty (z - \hat{c}) \bar{\Phi}(z) dz \\ &= \hat{c}^2 + 2 \int_0^\infty \left(z \bar{\Phi}(z) - \hat{c} \bar{\Phi}(z) \right) dz = \hat{c}^2 + \frac{1}{2} - \hat{c} \sqrt{\frac{2}{\pi}},\end{aligned}$$

where we bounded $\bar{\Phi}$ from above by 1. This case is particularly relevant when \mathcal{X} is a high-dimensional vector space: as the dimension of \mathcal{X} grows, the volume $\mu(C)$ of the corpus decreases. Moreover, as $\mu(C)$ tends to zero, then \hat{c} decreases to $-\infty$ and the tails of the conformance score are upper bounded by those of a Gaussian with mean converging to ∞ ; one then expects that the conformance concentrates to increasingly larger values as the dimension of \mathcal{X} grows. Thus, an increasingly larger portion of the data will achieve higher conformance scores. Anomalies should then be classified by adding \hat{c} to the conformance score.

A.2 One-class support vector machines

Here, we review one-class SVMs through the lens of the conditional value-at-risk (CVaR), a concept originating in financial risk management and stochastic programming, also known as the superquantile, expected shortfall, or average value-at-risk. The CVaR of a random variable Z summarises the average behaviour of the tail of its distribution. Takeda and Sugiyama (2008) showed that extended ν -support vector classifiers (E ν -SVC) introduced in (Pérez-Cruz et al., 2003) can be interpreted as minimising a CVaR. Subsequently, Tsyurmasto et al. (2014) showed that one-class SVMs can be also be interpreted as CVaR minimisation. More precisely, one-class SVMs seek $w \in \mathcal{H}$ that minimises the regularised CVaR

$$w^* = \arg \min_{w \in \mathcal{H}} \left\{ \text{CVaR}_\alpha(\langle w, \varphi(X) \rangle_{\mathcal{H}}) + \frac{1}{2} \|w\|_{\mathcal{H}}^2 \right\}.$$

Then, for any $\alpha \in [0, 1]$, the decision function is defined by $\mathbb{1}_{\Omega_\alpha^{\text{CVaR}}}(\cdot)$ where

$$\Omega_\alpha^{\text{CVaR}} := \{x \in \mathcal{X} : \langle w^*, \varphi(x) \rangle_{\mathcal{H}} \leq \text{VaR}_\alpha(\langle w^*, \varphi(X) \rangle_{\mathcal{H}})\}. \quad (\text{A.1})$$

By the variational formulation of the CVaR, we have

$$w^* = \arg \min_{w \in \mathcal{H}} \left\{ \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{\mathbb{E}_\mu[\langle w, \varphi(X) \rangle_{\mathcal{H}} - \eta]^+}{1 - \alpha} \right\} + \frac{1}{2} \|w\|_{\mathcal{H}}^2 \right\}.$$

The change of variables $\rho = -\eta$, $v = -w$, $\gamma = 1 - \alpha$ yields

$$v^* = \arg \min_{v \in \mathcal{H}} \left\{ \frac{1}{2} \|v\|_{\mathcal{H}}^2 + \min_{\rho \in \mathbb{R}} \left\{ \frac{1}{\gamma} \mathbb{E}_{\mu} [[-\langle v, \varphi(X) \rangle_{\mathcal{H}} + \rho]^+] - \rho \right\} \right\}.$$

When the measure μ is unknown, but a sample $x_1, \dots, x_n \sim \mu$ is available, an empirical estimator (Wang et al., 2014) may be formed

$$(\hat{v}^*, \hat{\rho}^*) = \arg \min_{v \in \mathcal{H}, \rho \in \mathbb{R}} \left\{ \frac{1}{2} \|v\|_{\mathcal{H}}^2 + \left(\frac{1}{\gamma n} \sum_{i=1}^n [\rho - \langle v, \varphi(x_i) \rangle_{\mathcal{H}}]^+ - \rho \right) \right\}, \quad (\text{A.2})$$

which is the unconstrained version of the quadratic programming problem (Xiao et al., 2017; Norton et al., 2017)

$$\begin{aligned} & \min_{v \in \mathcal{H}, \rho \in \mathbb{R}, \xi \in \mathbb{R}^n} \left\{ \frac{1}{2} \|v\|_{\mathcal{H}}^2 + \left(\frac{1}{\gamma n} \sum_{i=1}^n \xi_i - \rho \right) \right\}, \\ & \text{subject to } \langle v, \varphi(x_i) \rangle_{\mathcal{H}} \geq \rho - \xi_i \text{ and } \xi_i \geq 0, \text{ for all } i = 1, \dots, n \end{aligned}$$

whose solution is such that all training points are on one side of an hyperplane in the feature space with maximum margin. The empirical one-class SVM seeks a decision function of the form $\mathbf{1}_{\Omega_{\gamma}^{\text{ocsvm}}}(\cdot)$ (takes the value 1 for normal data and 0 otherwise where

$$\Omega_{\gamma}^{\text{ocsvm}} = \{x \in \mathcal{X} : \langle \hat{v}^*, \varphi(x) \rangle_{\mathcal{H}} \geq \hat{\rho}^*\}.$$

Appendix B. Conformance score for non-Gaussian streamed data

In practice, the assumption that the underlying sample comes from a Gaussian is restrictive. In the sequel we shall investigate under what assumptions on the distribution of the underlying data one can have similar types of estimates for large values of the conformance. The goal is to derive such guarantees in the setting of streamed data i.e. we assume that observations are points on the space $C([0, T]; \mathbb{R}^d)$.

Definition B.1 *The set of streams of d -dimensional data is defined as*

$$\mathcal{S}(\mathbb{R}^d) := \left\{ \mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n, n \in \mathbb{N} \right\}.$$

Given $k \in \mathbb{N}$ and a finite set $\{\mathbf{x}^i\}_{i=1}^k = \{(x_1^i, \dots, x_{n_i}^i)\}_{i=1}^k \subset \mathcal{S}(\mathbb{R}^d)$ of k streams, let $\ell(\mathbf{x}^i) := n_i$ denote the length of the i -th stream and

$$L := \max_{i=1, \dots, k} \ell(\mathbf{x}^i)$$

be the maximum stream length in the family $\{\mathbf{x}^i\}_{i=1}^k$. For each stream \mathbf{x}^i with $\ell(\mathbf{x}^i) < L$, consider the stream

$$\tilde{\mathbf{x}}^i = (x_1^i, x_2^i, \dots, x_{n_i}^i, x_{n_i}^i, \dots, x_{n_i}^i)$$

of length L augmented with the last observation in the remaining spots; note that the signature is invariant under this augmentation. Hence, the set $\{\tilde{\mathbf{x}}^i\}_{i=1}^k$ consists of streams of equal length L .

Assumption B.2 *The streams $\{\tilde{\mathbf{x}}^i\}_{i=1}^k$ are k realisations of an i.i.d. sample $X = (X^1, \dots, X^L)$ from an underlying distribution $\gamma \in \mathcal{P}(\mathbb{R}^d)$ with mean zero and a finite second moment i.e. for each $i = 1, \dots, k$, $X(\omega_i) = \tilde{\mathbf{x}}^i$ for some $\omega_i \in \mathbb{R}^d$. Moreover the covariance matrix $\Sigma^\gamma = \{\Sigma_{i,j}^\gamma\}_{i,j=1}^d := \{\mathbb{E}_\gamma[X_i^1 X_j^1]\}_{i,j=1}^d$ is invertible. A fortiori, we have*

$$\sigma_\gamma^2 := \text{tr} \Sigma^\gamma = \int_{\mathbb{R}^d} |x|_2^2 \gamma(dx) < \infty.$$

At this point, we shall transition to the view of data points as vectors on the path space $C([0, 1]; \mathbb{R}^d)$. Indeed, given an i.i.d. sample X from μ as above, we consider a continuous random path W^L defined via linear interpolation as

$$W_t^L := \frac{1}{\sqrt{L}} \sum_{k=1}^{[Lt]} X^k + \frac{(Lt - [Lt])}{\sqrt{L}} X^{[Lt]+1}, \quad \text{for all } t \in [0, 1]. \quad (\text{B.1})$$

Then, Condition B.2 placed above implies the following: *For each $i = 1, \dots, k$, the continuous path*

$$\phi^{i,L}(t) := \frac{1}{\sqrt{L}} \sum_{k=1}^{[Lt]} x_k^i + \frac{(Lt - [Lt])}{\sqrt{L}} x_{[Lt]+1}^i, \quad \text{for all } t \in [0, 1],$$

associated to the i -th stream $\tilde{\mathbf{x}}^i$, is a realisation of the random continuous path W^L .

Let us denote by μ^L the law of W^L on $C([0, 1]; \mathbb{R}^d)$. By the multidimensional version of Donsker's invariance principle (Revuz and Yor, 2013, Chapter XIII, Theorem 1.9), we know that as L tends to infinity, μ^L converges weakly in $C([0, 1]; \mathbb{R}^d)$ to the law of a d -dimensional Wiener process W with covariance

$$\mathbb{E}[W_t^i W_s^j] = \Sigma_{i,j}^\gamma(t \wedge s), \quad \text{for all } t, s \in [0, 1], i, j = 1, \dots, d. \quad (\text{B.2})$$

The following provides a guarantee for the topological conformance of a sample drawn from the approximating distribution:

Proposition B.3 *Let \mathcal{X} be a separable Banach space, C a Borel subset of \mathcal{X} and $\{\mu^L\}_{L \in \mathbb{N}} \in \mathcal{P}(\mathcal{X})$ a sequence of probability measures converging weakly to a Gaussian measure $\mu \in \mathcal{P}(\mathcal{X})$. If $K > 0$ is the smallest constant such that $\|x\|_{\mathcal{X}} \leq K \|x\|_\mu$ for all $x \in \mathcal{X}$ then, for all $r > 0$,*

$$\limsup_{L \rightarrow \infty} \mu^L \left(x \in \mathcal{X} : \text{dist}_{\mathcal{X}}(x, C) \geq r \right) \leq \bar{\Phi}(\hat{c} + r/K) \leq \exp \left\{ -\frac{(\hat{c} + r/K)^2}{2} \right\},$$

where $\hat{c} := \Phi^{-1}(\mu(C))$. Moreover, if C is closed, we can replace \hat{c} by $\Phi^{-1}(\limsup_{L \rightarrow \infty} \mu^L(C))$.

Proof Since $\mathcal{X} \ni x \mapsto \text{dist}_{\mathcal{X}}(x, C)$ is Lipschitz continuous, then $A := \{x : \text{dist}_{\mathcal{X}}(x, C) \geq r\}$ is a closed subset of \mathcal{X} . By the Portmanteau theorem for weak convergence of measures we know that $\limsup_{L \rightarrow \infty} \mu^L(A) \leq \mu(A)$. Putting these facts together we derive

$$\begin{aligned} \limsup_{L \rightarrow \infty} \mu^L \left(x \in \mathcal{X} : \text{dist}_{\mathcal{X}}(x, C) \geq r \right) &\leq \mu \left(x \in \mathcal{X} : \text{dist}_{\mathcal{X}}(x, C) \geq r \right) \\ &\leq \mu \left(x \in \mathcal{X} : \text{dist}_\gamma(x, C) \geq r/K \right) \\ &\leq \bar{\Phi}(\hat{c} + r/K). \end{aligned}$$

The last two inequalities are obtained using $\text{dist}_{\mathcal{X}}(x, C) \leq K \text{dist}_{\mu}(x, C)$ and the TSB inequality (Theorem A.4). The last assertion follows again from the Portmanteau theorem and the monotonicity of Φ^{-1} since $\bar{\Phi}(\hat{c} + r/K) \leq \bar{\Phi}(C + r/K)$ and

$$C := \Phi^{-1}\left(\limsup_{L \rightarrow \infty} \mu^L(C)\right) \leq \Phi^{-1}\left(\mu(C)\right) = \hat{c}.$$

■

Remark B.4 (Dimensionality considerations: the uncorrelated case) *The constant K in the theorem above does not depend on the dimension d of the data in the case where the covariance matrix Σ^γ is the identity. For example, if μ is the law of a standard d -dimensional Wiener process on $\mathcal{X} = C([0, 1]; \mathbb{R}^d)$, the corresponding Cameron-Martin space is given by*

$$\mathbb{H} := \left\{ x \in L^2([0, 1]; \mathbb{R}^d) : x(0) = 0, \dot{x} \in L^2([0, 1]; \mathbb{R}^d) \right\},$$

and for $x \in \mathbb{H}$,

$$\|x\|_{\mathcal{X}} := \sup_{t \in [0, 1]} |x(t)|_{\mathbb{R}^d} = \sup_{t \in [0, 1]} \left| \int_0^1 \dot{x}(t) dt \right|_{\mathbb{R}^d} \leq \left(\int_0^1 |\dot{x}(t)|_{\mathbb{R}^d}^2 dt \right)^{1/2} = \|x\|_{\mathbb{H}} \equiv \|x\|_{\mu}.$$

Hence K can be taken to be equal to 1. However, what does (implicitly) depend on dimension is the value of the topological conformance $\text{dist}_{\mathcal{X}}(x, C)$ because the norm of \mathcal{X} depends on d .

The following straightforward consequence of Proposition B.3 provides asymptotic conformance estimates in the particular case where μ is the law of a d -dimensional Wiener process with (invertible) covariance matrix Σ^γ .

Corollary B.5 *Let μ be the law of W in (B.2) and μ^L the law of the approximating path W^L in (B.1) on $C([0, 1]; \mathbb{R}^d)$. Under Condition B.2, for any Borel measurable subset $C \subset C([0, 1]; \mathbb{R}^d)$ and any $r > 0$,*

$$\begin{aligned} \limsup_{L \rightarrow \infty} \mu^L \left(x \in C([0, 1]; \mathbb{R}^d) : \text{dist}_{\mathcal{X}}(x, C) \geq r \right) &\leq \bar{\Phi} \left(\hat{c} + r \|\Sigma^\gamma\|_{op}^{\frac{1}{2}} \right) \\ &\leq \exp \left\{ -\frac{1}{2} \left(\hat{c} + r \|\Sigma^\gamma\|_{op}^{\frac{1}{2}} \right)^2 \right\}, \end{aligned} \tag{B.3}$$

where $\hat{c} := \Phi^{-1}(\mu(C))$ and $\|\cdot\|_{op}$ a matrix operator norm compatible with the norm on \mathbb{R}^d i.e. for all $x \in \mathbb{R}^d$, $|\Sigma^\gamma x|_{\mathbb{R}^d} \leq \|\Sigma^\gamma\|_{op} |x|_{\mathbb{R}^d}$. Moreover, if the set C is closed, we can replace the constant \hat{c} by $\Phi^{-1}(\limsup_{L \rightarrow \infty} \mu^L(C))$.

Proof By the multivariate version of Donsker's theorem (Revuz and Yor, 2013, Chapter XIII, Theorem 1.9), μ^L converges weakly to μ as $L \rightarrow \infty$ in the topology of $\mathcal{X} = C([0, 1]; \mathbb{R}^d)$. The variance/Cameron-Martin norm of μ is given by

$$\|x\|_{\mu}^2 = \int_0^1 |(\Sigma^\gamma)^{1/2} \dot{x}(t)|_{\mathbb{R}^d}^2 dt,$$

and

$$\|x\|_{\mathcal{X}} \leq \|\Sigma^\gamma\|_{op}^{-1/2} \left(\int_0^1 |(\Sigma^\gamma)^{1/2} \dot{x}(t)|_{\mathbb{R}^d}^2 dt \right)^{1/2} = \|\Sigma^\gamma\|_{op}^{-1/2} \|x\|_{\mu}.$$

This implies that the constant K in Proposition B.3 can be taken equal to $\|\Sigma^\gamma\|_{op}^{-1/2}$ for any matrix operator norm. \blacksquare

Corollary B.5 suggests that large values of the topological conformance $d_{\mathcal{X}}(x; C)$, under the measure μ^L , scale according to the covariance norm $\|\Sigma^\gamma\|_{op}$. As the dimension grows it is clear that this quantity increases. For example, if \mathbb{R}^d is equipped with the Euclidean (2-)norm, then $\|\Sigma^\gamma\|_{op}^2 = \|\Sigma^\gamma\|_2^2 = \max_{k=1, \dots, d} \lambda_k^2$, where λ_k are the eigenvalues of the matrix $\Sigma^\gamma(\Sigma^\gamma)^T$. Thus, in this case $\|\Sigma^\gamma\|_{op}$ is non-decreasing as a function of d .

In view of these facts, a sensible conformance measure of a stream $x \in C([0, 1]; \mathbb{R}^d)$, discretely sampled from an underlying distribution $\gamma \in \mathcal{P}(\mathbb{R}^d)$ with mean 0 and covariance Σ^γ , from a corpus $C \subset C([0, 1]; \mathbb{R}^d)$ would be the **variance-adjusted topological conformance**:

$$D_{\mathcal{X}, \gamma}(x; C) := \|\Sigma^\gamma\|_{op}^{1/2} dist_{\mathcal{X}}(x, C). \quad (\text{B.4})$$

Indeed, plugging the latter into (B.3) yields

$$\begin{aligned} \limsup_{L \rightarrow \infty} \mu^L \left(x \in C([0, 1]; \mathbb{R}^d) : D_{\mathcal{X}, \gamma}(x; C) \geq r \right) \\ = \limsup_{L \rightarrow \infty} \mu^L \left(x \in C([0, 1]; \mathbb{R}^d) : dist_{\mathcal{X}}(x; C) \geq \frac{r}{\|\Sigma^\gamma\|_{op}^{1/2}} \right) \leq \exp \left\{ -\frac{(\hat{c} + r)^2}{2} \right\}. \end{aligned}$$

The importance of this estimate lies in the fact that probabilities of high conformance scores are not affected by dimensionality as long as the scores are appropriately adjusted.

Appendix C. Experimental Details

The raw nanopore direct sequencing data were obtained from the European Nucleotide Archive (accession number PRJEB44511). Specifically we downloaded the modified and unmodified samples with respective experiment accession numbers ERX6983963 and ERX6983965. We converted the samples into the pod5 format using the command

```
pod5 convert fast5 <path_to_fast5>/*.fast5 --output Oligo_<Mod_or_Unmod>.pod5
```

Using the (.fa) reference sequence

```
>control
ATACTCGACATAGATAGGACTCTTTAGCTAGTGAACCCTAGCCTCCGGAGACAGGTCGCGACCTGTGTAGATGAGAGAAC
TGAGTGCACAAAAAAAAAAAA
```

we basecalled the sample using Dorado (version 0.9.1) using the rna002_70bps_hac@v3 model using the command

```
dorado basecaller <path_to_dorado_model>/rna002_70bps_hac@v3 sample.pod5 \
--reference ref.fa --emit-moves --emit-sam > basecalled_sample.bam
```

and event-level signal segmentation was obtained with Uncalled4, running the command

```
uncalled4 align --ref ref.fa --reads sample.pod5 --bam-in basecalled_sample.bam \
--eventalign-out --flowcell FLO-MIN106 --kit SQK-RNA002 --min-aln-length 5 \
--eventalign-flags print-read-names,signal-index,samples > aligned_sample.txt
```

For the modified sample, we had to additionally run the following commands before signal alignment with Uncalled4

```
samtools fastq -@8 -T "mv,ts,pi,sp,ns" basecalled_sample.bam > sample.moves.fastq
minimap2 -y -ax map-ont -k9 -w1 --secondary=no -t8 ref.fa samples.moves.fastq \
| samtools sort -@8 -o sample.mm2.bam
```

References

- E. Abi Jaber and L.-A. Gérard. Signature volatility models: pricing and hedging with Fourier. *SIAM Journal on Financial Mathematics*, 16(2):606–642, 2025.
- H. Andrès, A. Boumezoued, and B. Jourdain. Signature-based validation of real-world economic scenarios. *ASTIN Bulletin: The Journal of the IAA*, 54(2):410–440, 2024.
- I. P. Arribas. Derivatives pricing using signature payoffs. *arXiv:1809.09466*, 2018.
- I. P. Arribas, C. Salvi, and L. Szpruch. Sig-SDEs model for quantitative finance. In *ACM International Conference on AI in Finance*, 2020.
- P. Arrubarrena, M. Lemerrier, B. Nikolic, T. Lyons, and T. Cass. Novelty detection on radio astronomy data using signatures. *arXiv:2402.14892*, 2024.
- B. Barancikova, Z. Huang, and C. Salvi. Sigdiffusions: Score-based diffusion models for time series via log-signature embeddings. *ICLR*, 2025.

- S. Bates, E. Candès, L. Lei, Y. Romano, and M. Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178, 2023.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical Society: Series B*, 57(1): 289–300, 1995.
- H. Boedihardjo. Decay rate of iterated integrals of branched rough paths. *Annales de l’IHP: Analyse Non Linéaire*, 35(4):945–969, 2018.
- O. Bonesini, E. Ferrucci, I. Gasteratos, and A. Jacquier. Rough differential equations for volatility. *arXiv:2412.21192*, 2024.
- R. Bouman, Z. Bukhsh, and T. Heskes. Unsupervised anomaly detection algorithms on real-world data: how many do we need? *Journal of Machine Learning Research*, 25(105): 1–34, 2024.
- H. Buehler, B. Horvath, T. Lyons, I. P. Arribas, and B. Wood. A data-driven market simulator for small data environments. *arXiv:2006.14498*, 2020.
- T. Cass and C. Salvi. Lecture Notes on Rough Paths and Applications to Machine Learning. *arXiv:2404.06583*, 2024.
- T. Cass, C. Litterer, and T. J. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *The Annals of Probability*, 41(4):3026–3050, 2013.
- T. Cass, L. Gonon, and N. Zozoulenko. Variance norms for kernelized anomaly detection. *arXiv:2407.11873*, 2024.
- C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–27, 2011.
- I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, 23(176):1–42, 2022.
- T. Cochrane, P. Foster, T. Lyons, and I. P. Arribas. Anomaly detection on streamed data. *arXiv:2006.03487*, 2020.
- T. Cochrane, P. Foster, V. Chhabra, M. Lemerrier, T. Lyons, and C. Salvi. SK-Tree: a systematic malware detection algorithm on streaming trees via the signature kernel. In *IEEE International Conference on Cyber Security and Resilience*, pages 35–40, 2021.
- S. Crew, C. Salvi, W. F. Turner, T. Cass, and A. Jacquier. Quantum path signatures. *arXiv:2508.05103*, 2025.
- C. Cuchiero and J. Möller. Signature methods in stochastic portfolio theory. *arXiv:2310.02322*, 2023.
- B. Davis and I. Monroe. Randomly started signals with white noise. *The Annals of Probability*, pages 922–925, 1984.

- B. Davis and E. Perkins. Brownian slow points: the critical case. *The Annals of Probability*, pages 779–803, 1985.
- P. Díaz Lozano, T. Lozano Bagén, and J. Vives. Neural stochastic differential equations for conditional time series generation using the Signature Wasserstein-1 metric. *Journal of Computational Finance*, 27:1–23, 2023.
- A. Fermanian, T. Lyons, J. Morrill, and C. Salvi. New directions in the applications of rough path theory. *IEEE BITS the Information Theory Magazine*, 3(2):41–53, 2023.
- P. K. Friz and M. Hairer. *A Course on Rough Paths*. Springer, 2020.
- M. Furlan, A. Delgado-Tejedor, L. Mulroney, M. Pelizzola, E. M. Novoa, and T. Leonardi. Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biology*, 18(sup1):31–40, 2021.
- I. Gasteratos and A. Jacquier. Transportation-cost inequalities for non-linear Gaussian functionals. *arXiv:2310.05750*, 2023.
- N. Gozlan and C. Léonard. Transport inequalities. a survey. *Markov Processes And Related Fields*, 16:635–736, 2010.
- M. Hairer. An introduction to stochastic PDEs. *arXiv:0907.4178*, 2009.
- D. Hitchcock. *Some Geometric Approaches to Parameter Estimation*. PhD thesis, University of Edinburgh, 1992.
- M. Hoglund, E. Ferrucci, C. Hernández, A. M. Gonzalez, C. Salvi, L. Sánchez-Betancourt, and Y. Zhang. A neural RDE approach for continuous-time non-Markovian stochastic control problems. *ICML*, 2023.
- C. Holberg and C. Salvi. Exact gradients for stochastic spiking neural networks driven by rough signals. *NeurIPS*, 37:31907–31939, 2024.
- B. Horvath, M. Lemerrier, C. Liu, T. Lyons, and C. Salvi. Optimal stopping via distribution regression: a higher rank signature approach. *arXiv:2304.01479*, 2023.
- Z. Issa, B. Horvath, M. Lemerrier, and C. Salvi. Non-adversarial training of neural SDEs with signature kernel scores. *NeurIPS*, 36:11102–11126, 2023.
- P. Kidger and T. Lyons. Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU. *arXiv:2001.00706*, 2021. URL <https://github.com/patrick-kidger/signatory>.
- P. Kidger, P. Bonnier, I. Perez Arribas, C. Salvi, and T. Lyons. Deep signature transforms. *NeurIPS*, 32, 2019.
- F. J. Király and H. Oberhauser. Kernels for sequentially ordered data. *Journal of Machine Learning Research*, 20(31):1–45, 2019.

- S. Kovaka, P. W. Hook, K. M. Jenike, V. Shivakumar, L. B. Morina, R. Razaghi, W. Timp, and M. C. Schatz. Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nature Methods*, pages 1–11, 2025.
- A. Leger, P. P. Amaral, L. Pandolfini, C. Capitanchik, F. Capraro, V. Miano, V. Migliori, P. Toolan-Kerr, T. Sideri, A. J. Enright, et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nature communications*, 12(1):7198, 2021.
- M. Lemerrier, C. Salvi, T. Damoulas, E. Bonilla, and T. Lyons. Distribution regression for sequential data. In *International Conference on Artificial Intelligence and Statistics*, pages 3754–3762. PMLR, 2021.
- M. Lemerrier, P. Arrubarrrena, S. Di Giorgio, J. Brettschneider, T. Cass, I. S. Vries, A. Papavasiliou, A. Ruggieri, I. Tellioglu, C. C. Wu, et al. Path signatures enable model-free mapping of rna modifications. *arXiv preprint arXiv:2511.08855*, 2025.
- G. Li and J. J. Jung. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91:93–102, 2023.
- H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, 2018.
- T. Lyons and al. CoRoPa: Computational rough paths (software library), 2010. URL <https://coropa.sourceforge.io/>.
- T. Lyons and A. McLeod. Signature methods in machine learning. *EMS Surveys in Mathematical Sciences*, 2025.
- G. Manten, C. Casolo, E. Ferrucci, S. W. Mogensen, C. Salvi, and N. Kilbertus. Signature kernel conditional independence tests in causal discovery for stochastic processes. *ICLR*, 2025.
- R. Mignot, V. Mangé, K. Usevich, M. Clausel, J.-Y. Tournier, and F. Vincent. Anomaly detection using multiscale signatures. In *32nd European Signal Processing Conference (EUSIPCO)*, pages 2757–2761, 2024.
- J. Morrill, C. Salvi, P. Kidger, and J. Foster. Neural rough differential equations for long time series. *ICML*, pages 7829–7838, 2021.
- G. Muñoz Gil, G. G. i Corominas, and M. Lewenstein. Unsupervised learning of anomalous diffusion data: an anomaly detection approach. *Journal of Physics A: Mathematical and Theoretical*, 54(50):504001, 2021.
- N. Muça Cirone and C. Salvi. Parallelfow: Parallelizing linear transformers via flow discretization. *arXiv:2504.00492*, 2025a.
- N. Muça Cirone and C. Salvi. Rough kernel hedging. *arXiv:2501.09683*, 2025b.
- N. Muça Cirone, M. Lemerrier, and C. Salvi. Neural signature kernels as infinite-width-depth-limits of controlled resnets. *ICML*, pages 25358–25425, 2023.

- N. Mua Cirone, A. Orvieto, B. Walker, C. Salvi, and T. Lyons. Theoretical foundations of deep selective state-space models. *arXiv:2402.19047*, 2024.
- M. Norton, A. Mafusalov, and S. Uryasev. Soft margin support vector classification as buffered probability minimization. *Journal of Machine Learning Research*, 18(68):1–43, 2017.
- A. Pannier and C. Salvi. A path-dependent PDE solver based on signature kernels. *arXiv:2403.11738*, 2024.
- I. Perez Arribas. *Signatures in Machine Learning and Finance*. PhD thesis, University of Oxford, 2020.
- F. P  rez-Cruz, J. Weston, D. Herrmann, and B. Sch  lkopf. Extension of the ν -SVM range for classification. In *Advances in Learning Theory: Methods, Models and Applications*, volume 190, pages 179–196. IOS Press, 2003.
- J. Reizenstein and B. Graham. The iisignature library: efficient calculation of iterated-integral signatures and log signatures. *arXiv:1802.08252*, 2018.
- D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293. Springer, 2013.
- R. T. Rockafellar and S. Uryasev. Conditional Value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.
- L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. M  ller, and M. Kloft. Deep one-class classification. *ICML*, pages 4393–4402, 2018.
- L. Ruff, Y. Zemlyanskiy, R. Vandermeulen, T. Schnake, and M. Kloft. Self-attentive, multi-context one-class classification for unsupervised anomaly detection on text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4061–4071, 2019.
- C. Salvi. *Rough Paths, Kernels, Differential Equations and an Algebra of Functions on Streams*. PhD thesis, University of Oxford, 2021.
- C. Salvi, T. Cass, J. Foster, T. Lyons, and W. Yang. The Signature Kernel is the solution of a Goursat PDE. *SIAM Journal on Mathematics of Data Science*, 3(3):873–899, 2021a.
- C. Salvi, M. Lemerrier, C. Liu, B. Horvath, T. Damoulas, and T. Lyons. Higher order kernel mean embeddings to capture filtrations of stochastic processes. *NeurIPS*, 34:16635–16647, 2021b.
- C. Salvi, M. Lemerrier, and A. Gerasimovics. Neural stochastic PDEs: Resolution-invariant learning of continuous spatiotemporal dynamics. *NeurIPS*, 35:1333–1344, 2022.
- C. Salvi, J. Diehl, T. Lyons, R. Preiss, and J. Reizenstein. A structure theorem for streamed information. *Journal of Algebra*, 634:911–938, 2023.

- B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.
- R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 2009.
- Z. Shao, R. Sze-Yin Chan, T. Cochrane, P. Foster, and T. Lyons. Dimensionless anomaly detection on multivariate streams with variance norm and path signature. *arXiv:2006.03487*, 2020.
- D. Shmelev and C. Salvi. Sparse signature coefficient recovery via kernels. *arXiv:2412.08579*, 2024.
- D. Shmelev and C. Salvi. pySigLib-Fast Signature-Based Computations on CPU and GPU. *arXiv:2509.10613*, 2025.
- G. Sikora, K. Burnecki, and A. Wylomańska. Mean-squared-displacement statistical test for fractional Brownian motion. *Physical Review E*, 95(3):032110, 2017.
- G. Staerman, P. Mozharovskyi, S. Cléménçon, and F. d’Alché Buc. Functional isolation forest. *Asian Conference on Machine Learning*, pages 332–347, 2019.
- G. Staerman, M. Campi, and G. W. Peters. Signature isolation forest. *arXiv:2403.04405*, 2024.
- A. Takeda and M. Sugiyama. ν -support vector machine as conditional Value-at-risk minimization. *ICML*, pages 1056–1063, 2008.
- M. Talagrand. Transportation cost for Gaussian and other product measures. *Geometric & Functional Analysis GAFA*, 6(3):587–600, 1996.
- P. Tsyurmasto, M. Zabarankin, and S. Uryasev. Value-at-risk support vector machine: stability to outliers. *Journal of Combinatorial Optimization*, 28(1):218–232, 2014.
- B. Walker, L. Yang, N. Muça Cirone, C. Salvi, and T. Lyons. Structured linear CDEs: Maximally expressive and parallel-in-time sequence models. *arXiv:2505.17761*, 2025.
- Y. Wang, C. Dang, and S. Wang. Robust novelty detection via worst case CVaR minimization. *IEEE Transactions on Neural Networks and Learning Systems*, 26(9):2098–2110, 2014.
- L. K. White and J. R. Hesselberth. Modification mapping by nanopore sequencing. *Frontiers in Genetics*, 13:1037134, 2022.
- M. J. Wilensky, M. F. Morales, B. J. Hazelton, N. Barry, R. Byrne, and S. Roy. Absolving the SSINS of precision interferometric radio data: a new technique for mitigating faint radio frequency interference. *Publications of the Astronomical Society of the Pacific*, 131(1005):114507, 2019.

- G. Wynne and A. B. Duncan. A kernel two-sample test for functional data. *Journal of Machine Learning Research*, 23(73):1–51, 2022.
- Y. Xiao, H. Wang, and W. Xu. Ramp loss based robust one-class SVM. *Pattern Recognition Letters*, 85:15–20, 2017.
- Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi. Deep learning for time series anomaly detection: a survey. *ACM Computing Surveys*, 57(1):1–42, 2024.