# Anomaly Detection Using Multiscale Signatures

Raphael MIGNOT
*IECL*
*University of Lorraine*
Nancy, France
raphael.mignot@univ-lorraine.fr

Valérian MANGÉ
*Hensoldt Nexeya France*[1]
*& TéSA Laboratory*[2]
Toulouse, France
[1]valerian.mange@hensoldt.fr
[2]valerian.mange@tesa.prd.fr

Konstantin USEVICH
*CRAN*
*University of Lorraine*
Nancy, France
konstantin.usevich@univ-lorraine.fr

Marianne CLAUSEL
*IECL*
*University of Lorraine*
Nancy, France
marianne.clausel@univ-lorraine.fr

Jean-Yves TOURNERET
*University of Toulouse, ENSEEIHT-IRIT*
*& TéSA Laboratory*
Toulouse, France
jean-yves.tourneret@toulouse-inp.fr

François VINCENT
*ISAE-Supaéro*
Toulouse, France
francois.vincent@isae-supaero.fr

***Abstract*—This paper analyzes multidimensional time series through the lens of their integrals of various moment orders, constituting their signatures, a novel tool for detecting anomalies in time series. The proposed anomaly detection (AD) method is compared using classical distance-based methods such as Local Outlier Factor (LOF) and One-Class Support Vector Machine (OCSVM). These methods are investigated using different similarity measures: distance on signature features, Euclidean distance and Dynamic Time Warping (DTW). The combination of signature features with a specific segmentation of time series leads to a multi-scale analysis tool that is competitive with respect to the state-of-the-art results, while maintaining low computational costs thanks to a property of the signature features.**

***Index Terms*—Time series, Anomaly detection, Iterated integrals signature, Segmentation, Unsupervised learning.**

## I. Introduction

**Context.** Anomaly detection (AD) is a critical field of research with various applications in different fields (medical, telemetry, etc.) [1], [2]. In certain fields, the data to be analyzed is a set of time series. Along with the typical challenges of AD, anomalies in time series can appear in multiple forms such as global anomalies or contextual anomalies [2]. This raises the need for algorithms that can operate at multiple scales in order to identify the nature of anomalies and determine their exact location with a reasonable time complexity. Furthermore, specific challenges are related to the analysis of time series (irregularly sampled data, missing measurements, different recording lengths, etc.). These issues make the problem of AD in time series an active research field.

**State-of-the-art.** Several AD algorithms exist in the state-of-the-art [3], such as Isolation Forest [4], Local Outlier Factor (LOF) [5] and One-Class SVM (OCSVM) [6]. These algorithms are well suited for comparing vectors in a space of fixed dimension. To compare time series with different lengths, similarity measures such as the Dynamic Time Warping (DTW) have been proposed in the literature [7]. These techniques have already been incorporated in the standard AD algorithms for time series [8]. However, they need to compute all the pairwise similarities, which is time consuming.

This paper proposes to analyze time series through the lens of their signature features, allowing dependencies among components of multivariate time series to be detected. Signature features have been originally developed for topological work [9] and later on have been used in the theory of Rough Paths [10]. Due to their ability to encode non linear dependencies in multivariate time series, signature features have been successfully used to address numerous ML task, see [11]. In this context, this novel set of features has provided state-of-the-art performances in a number of applications such as handwriting recognition [12], medical condition detection [13] and oceanography [14].

**Objectives, contributions and organization.** An original approach for AD in time series is introduced in this paper using standard AD algorithms and multi-scale signatures features. The approach is compared to multiple state-of-the-art algorithms in terms of detection performance and time complexity. The contributions of this work are summarized below:

- We show that a multi-scale analysis using signature features leads to state-of-the-art results for AD in multivariate time series.
- In addition to excellent detection results, we put in evidence the incredible numerical effectiveness of the multiscale signature based analysis. Notably, we avoid huge computational burden, that would appear with other similarity measures. The signature multiscale method can be 100 times faster than single-scale DTW.

Section II recalls the principles of LOF and OCSVM methods that are used in this paper. Section III provides details on DTW and signature methods. Section IV compares the proposed methods with state-of-the-art approaches on both synthetic and real datasets. Conclusions are reported in Section V.

## II. Anomaly Detection

The objective of AD is to detect abnormal behavior, i.e., data that deviates significantly from what is observed in the majority of cases. Abnormalities can represent different phenomena depending on the data that is analyzed. In general, abnormal data is scarce and cannot be used to describe all possible anomalies [3]. Therefore, a model must be trained in an unsupervised way while considering an imbalance between the normal and abnormal classes. State-of-the-art AD algorithms include LOF and OCSVM that are considered in this paper (other algorithms such as Isolation Forest and Density Based Spatial Clustering of Applications with Noise (DBSCAN) could be considered similarly).

### A. Local Outlier Factor (LOF)

LOF [5] is an AD algorithm based on the density of the training data for classification. The more abnormal a data vector is, the larger its distance to its neighbors. Consider $N \geq 1$ vectors $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ with $\mathbf{x}_n \in \mathbb{R}^D$, $n = 1, \dots, N$, $D$ being the dimension of the vectors $\mathbf{x}_n$ and $k$ a strictly positive integer. The distance between $\mathbf{x}_n$ and its $k$-th nearest neighbor is referred to as $k$-distance of $\mathbf{x}_n$ and denoted as $\mathrm{kd}(\mathbf{x}_n)$. The reachability distance between $\mathbf{x}_n$ and $\mathbf{x}_m$ is $\mathrm{rd}_k(\mathbf{x}_n, \mathbf{x}_m) = \max\{\mathrm{kd}(\mathbf{x}_n), \|\mathbf{x}_n - \mathbf{x}_m\|\}$. Therefore, all the points in the neighborhood of $\mathbf{x}_n$ have the same reachability but points that are further away will have a higher reachability. The local reachability density of $\mathbf{x}_n$ is the inverse of the average of all reachability distances of the $k$-nearest neighbors of $\mathbf{x}_n$: $\mathrm{lrd}_k(\mathbf{x}_n) = k \left( \sum_{\mathbf{x} \in V_n}^{N} \mathrm{rd}_k(\mathbf{x}, \mathbf{x}_n) \right)^{-1}$ with $V_n$ the subset containing all $k$-nearest neighbors of $\mathbf{x}_n$. The LOF score of $\mathbf{x}_n$ is the mean ratio of the local reachability densities of $\mathbf{x}_n$ with its $k$-nearest neighbors: $\mathrm{LOF}_k(\mathbf{x}_n) = \frac{1}{k} \sum_{\mathbf{x} \in V_n} \frac{\mathrm{lrd}_k(\mathbf{x})}{\mathrm{lrd}_k(\mathbf{x}_n)}$. Inliers have a LOF score close to 1 whereas outliers have a much higher LOF score.

### B. One-Class SVM

*1) Method:* One-Class SVM is an AD algorithm, which learns a separating hyperplane close to the normal data in a certain sense. In its basic form, the data must be linearly separable in their vector space of the data. Consider a training set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ with $N \geq 1$ and $\forall i \in \{1, \dots, N\}$, $\mathbf{x}_i \in \mathbb{R}^D$ with $D$ the number of data features. Finding a hyperplane separating the data can be expressed as

$$\begin{aligned} \underset{\omega, \rho, \xi_i}{\text{minimize}} \quad & \frac{1}{2}\|\omega\|^2 - \rho + \frac{1}{\nu N} \sum_{i=1}^{N} \xi_i \\ \text{with} \quad & \langle \omega, \mathbf{x}_i \rangle \geq \rho - \xi_i, \forall i \in \{1, \dots, N\} \end{aligned} \tag{1}$$

where $\langle \omega, \mathbf{x}_i \rangle = \omega^T \mathbf{x}_i$ and $\nu$ is the maximum proportion of abnormal data in $\mathcal{X}$.

*2) Kernel trick:* When the data is not linearly separable, a reproducing kernel $\kappa$ can be used. This kernel is associated with a scalar product in a higher dimensional space such that $\kappa(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$ and $\forall i \in \{1, \dots, N\}, \Phi(\mathbf{x}_i) \in \mathbb{R}^F$ with $F \geq D$. The kernel function should be chosen in a way that the normal data is linearly separable from anomalies in the new space. The constraints in (1) are now applied to the transformed vectors:

$$\langle \omega, \Phi(\mathbf{x}_i) \rangle \geq \rho - \xi_i, \forall i \in \{1, \dots, N\} \tag{2}$$

with the new decision function $f(\mathbf{x}) = \mathrm{sgn}(\langle \omega, \Phi(\mathbf{x}) \rangle - \rho)$ where sgn denotes the "sign" function. The solution of (1) with the constraints (2) is known to be $\omega = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i)$, $0 \leq \alpha_i \leq \frac{1}{\nu N}$ s.t. $\sum_{i=1}^{N} \alpha_i = 1$ and the decision function is $f(\mathbf{x}) = \mathrm{sgn}\left( \sum_{i=1}^{N} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) - \rho \right)$. A common choice is the Gaussian kernel (or radial basis function) defined by $\kappa(\mathbf{x}, \mathbf{y}) = \exp\left( -\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2} \right)$ where $\sigma$ is a hyper-parameter to be estimated. It can be shown that the corresponding function $\Phi$ projects the data into a space of infinite dimension.

### C. Performance evaluation

Appropriate metrics must be defined to evaluate the performances of AD accurately. The measures of precision, recall and a combination of the two, the $\mathrm{F}_1$ score, are usually considered for AD [15]. These measures are defined as:

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \ \text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives. The $\mathrm{F}_1$ score is the harmonic mean of precision and recall. Precision-Recall curves are often preferred to ROC curves for AD [15].

## III. Comparing trajectories

### A. Feature engineering: Signatures

The signature method can be understood as a time series version of the method of moments for random variables [16]. Roughly speaking, signature features encode high order dependencies between the components of a multivariate time series. Formally, given a continuous function of finite total variation $X : [0, T] \to \mathbb{R}^d$, with $T$ a positive real value, then the signature of level (or order) $L$ of $X$ is a tensor in $(\mathbb{R}^d)^{\otimes L}$ defined as

$$\mathbf{S}_{(L)}(X) = \int_{0 \leq t_1 \leq \dots \leq t_L \leq T} \dots \int \dot{X}(t_1) \otimes \dots \otimes \dot{X}(t_L) dt_1 \dots dt_L \tag{4}$$

where $\otimes$ is the tensor product and $\dot{X}$ is the derivative of $X$. The signature of $X$ is the infinite collection of signatures at all levels:

$$\mathbf{S}(X) = \{\mathbf{S}_{(1)}(X), \mathbf{S}_{(2)}(X), \dots\}. \tag{5}$$

As given by (4), $\mathbf{S}_{(1)}(X)$ (level $L = 1$) is the global change $X(T) - X(0)$, $\mathbf{S}_{(2)}(X)$ (level $L = 2$) is a matrix whose coefficients are related to correlations and cross correlations of the stochastic process $X$ [17]. In numerical experiments, the computation of the signature is done up to a fixed level $L$. The collection of the first $L$ elements of $\mathbf{S}(X)$ is denoted as $\mathbf{S}_{\leq L}(X)$. Signatures can be useful for AD for multiple reasons:

- $\mathbf{S}(X)$ uniquely determines $X$ up to translation and time reparametrization, except for pathological cases [18]. The

invariance regarding time parametrization is in the following sense: given a non decreasing continuous surjection $\varphi$ on $[0,T]$, denote as $\tilde{X} : [0,T] \to \mathbb{R}^d$; $t \mapsto X(\varphi(t))$. Then, $\mathbf{S}(X) = \mathbf{S}(\tilde{X})$. Thus the signature function $\mathbf{S}$ can be seen as a non linear filter, providing features invariant to translation and resampling.

- The dimension of $\mathbf{S}_{\leq L}(X)$ does not depend on the length $T$. Thus the signatures of two signals of different lengths $T_1$ and $T_2$ can be easily compared. Indeed, if one vectorizes the signatures at different levels, the associated signature vector, denoted as $\mathrm{vec}(\mathbf{S}_{\leq L}(X))$, is of dimension $\sum_{k=1}^{L} d^k = \frac{d(d^L-1)}{d-1}$.

Note that any non linear function of a multivariate time series can be arbitrarily well approximated by a linear function of its signature.

To illustrate the true meaning of (4), consider as an example the function $X : t \in [0,T] \mapsto (X_1(t), \ldots, X_d(t))$ defined as an affine trajectory, i.e., $X(t) = at + b$, with $a, b \in \mathbb{R}^d$. To compute the value of the signature of $X$ at level $L \in \mathbb{N}$, one can apply (4) and find the expression of the corresponding $L$-ways tensor at any set of indices $1 \leq i_1, \ldots, i_L \leq d$,

$$\left[\mathbf{S}_{(L)}(X)\right]_{i_1,\ldots,i_L} = \frac{1}{L!} \prod_{k=1}^{L} (X_{i_k}(T) - X_{i_k}(0)) \in \mathbb{R}. \quad (6)$$

A similarity measure $d_{\mathrm{sig}}$ is defined by computing the Euclidean distance between two vectorized signatures, i.e., $d_{\mathrm{sig}}(X,Y) = \|\mathrm{vec}(\mathbf{S}_{\leq L}(X)) - \mathrm{vec}(\mathbf{S}_{\leq L}(Y))\|_2$, for two trajectories $X$ and $Y$. Note that in some cases, normalizing the coefficients of the signature can improve AD performances. This normalization can be performed, e.g., by replacing $\mathbf{S}_{(k)}(X)$ with $(\mathbf{S}_{(k)}(X))^{1/k}$ where the exponent is applied element-wise, for any $k = 1, \ldots, L$.

Note that from a computational point of view, the computation of the signature is very simple. A fundamental algebraic property of the signature is the following: the space of signature features forms a Lie group and the corresponding group operation is related to the concatenation of trajectories. For two functions $X : [0,u] \to \mathbb{R}^d$ and $Y : [u,T] \to \mathbb{R}^d$, $0 < u < T$, one has

$$\mathbf{S}_{[0,T]}(X \star Y) = \mathbf{S}_{[0,u]}(X) \cdot \mathbf{S}_{[u,T]}(Y), \quad (7)$$

where $\mathbf{S}_{[a,b]}$ is the iterated integrals as in (4) but with lower and upper bounds $a$ and $b$ (instead of 0 and $T$), $\star$ denotes concatenation and $\cdot$ is the multiplication operation between two signatures. Fig. 1 illustrates (7). As a consequence, one can compute the signature of a time series combining a piecewise linear approximation and this last property. More precisely, in order to handle time series (discrete set of points) a linear interpolation can be performed before calculating the signature. Because (7) depends on the concatenation operation, it is affected by the interpolation method and the linear interpolation leads to efficient numerical computations. Thereafter, for any time series $X$, $\mathbf{S}(X)$ denotes the signature of the linearly interpolated time series. Note that the signature has been used for AD in a semi-supervised context in [19] where

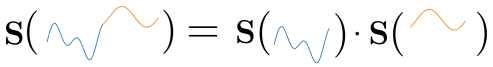


Fig. 1: Illustration of (7) with a one dimensional trajectory. $X$ is represented by the blue curve and $Y$ by the orange curve.

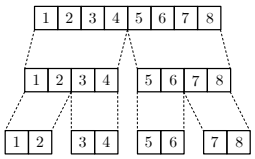



Fig. 2: Dyadic windowing operation on a time series of length $n = 8$. We obtain the three rows by applying successively $W^1, W^2$ and $W^3$ to $X$.

an AD algorithm called `SigMahaKNN` (SMK) was based on anomaly scores using a generalized Mahalanobis distance on the signature features. This paper proposes to detect anomalies in trajectories using signature features with a different and unsupervised approach, as explained below.

### *B. Multi-scale signature feature*

A powerful tool which allows for multiscale analysis is the hierarchical dyadic windowing operation defined in what follows. This operation is interesting in the case of signature features for two reasons: it has shown to improve classification results [20] and it does not add computational overheads as explained below. Let $\{X(t_i)\}_{i=1,\ldots,n}$ be a multivariate time series. Denote $\ell \in \mathbb{N}$ the hierarchical depth. For simplicity, assume that $2^{\ell-1}$ divides $n$. Denote as $W^i(X)$ the time series obtained after segmentation with a sliding window of length equal to $\frac{n}{2^{(i-1)}}$, for $i = 1, \ldots, \ell$, as shown in Fig. 2. Denote as $\mathrm{MSIG}(L) = \mathbf{S}_{\leq L} \circ W$ the signature computed on the segmented time series. The multi-scale signature method for AD is defined by:

$$\mathrm{AD} \circ \mathrm{MSIG}(L). \quad (8)$$

The windowing operation $W$ in MSIG does not add computational overheads since operations $\mathbf{S}_{\leq L}(X)$ and $(\mathbf{S}_{\leq L} \circ W)(X)$ have the same complexity. Indeed, the computation of the signature of $X$ is done in two steps: 1) apply (6) on the $n/2$ pairs of the form $\{X(t_{2k}), X(t_{2k+1})\}$, 2) perform $\ell$ successive iterations where signatures are combined using (7). These two steps are illustrated in Fig. 2 read from bottom to top: in the first iteration, pairs are combined by computing the product of the signatures of $\{X(t_1), X(t_2)\}$ and $\{X(t_3), X(t_4)\}$ using (7), which gives the signature of the time series $\{X(t_1), X(t_2), X(t_3), X(t_4)\}$. At the end of this process, the signature of the whole time series $\{X(t_i)\}_{i=1,\ldots,n}$ is obtained. Therefore, to obtain $(\mathbf{S} \circ W)(X)$, it suffices to store all the signatures of sub time series calculated during the computation process of $\mathbf{S}(X)$.

Note that each time we go deeper in the dyadic segmentation, the more of the resampling invariance is lost.

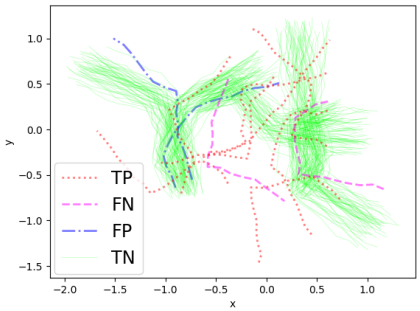



Fig. 3: Examples of synthetic trajectories after prediction using LOF with SIG4 (2 FPs and 2 FNs).

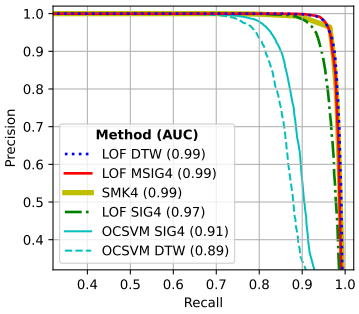



Fig. 4: Precision-recall curves using LOF and OCSVM with various similarity measures.

### C. Trajectory alignment : Dynamic Time Warping

Another way to compare two trajectories directly is to find which points of one trajectory *best* match the points of the other trajectory. This technique referred to as trajectory alignment requires to find associations or removes points from the two trajectories. A common tool used for trajectory alignment is DTW [21]. Consider $X = (X(1), \ldots, X(n))$ and $Y = (Y(1), \ldots, Y(m))$ two $d$-dimensional time series, i.e., $X(i) \in \mathbb{R}^d$ and $Y(j) \in \mathbb{R}^d$. The DTW cumulated similarity score between the first $i$ columns of $X$ and the first $j$ columns of $Y$, denoted as $\mathrm{cs}(i, j)$, is defined as

$$\mathrm{cs}(i,j) = \min\{\mathrm{cs}(i-1,j-1), \mathrm{cs}(i,j-1), \mathrm{cs}(i-1,j)\} + \|X(i) - Y(j)\|^2. \quad (9)$$

## IV. Experiments

The proposed AD methods are based on computing signatures and applying the LOF or One-Class SVM methods with the distance $d_{\mathrm{sig}}$. Two kinds of signatures are considered yielding the algorithms SIG4 (using $\mathbf{S}_{\leq L}(X)$ with $L = 4$) and MSIG4 (using $(\mathbf{S}_{\leq L} \circ W)(X)$ with $L = 4$). These methods are compared to DTW combined with LOF and DTW with One-Class SVM (OCSVM) to show the interest of using signatures, and to another AD algorithm `SigMahaKNN` (SMK). The number of neighbors in LOF was set to 9. Experiments can be reproduced using the notebook at https://github.com/Raph-AI/anomaly-detection.

### A. Synthetic data

The first set of experiments consists in analyzing synthetic data representing trajectories in the context of abnormal ship behavior [22]. The dataset contains $N_T = 1000$ sets of trajectories, each with 260 time series, 10 of them being abnormal according to the generated ground truth. The remaining 250 trajectories are divided into 5 groups (or railways) of 50 trajectories. The trajectories are each composed of 16 positions in a 2D space. Examples of trajectories are displayed in Fig. 3. The method will not be compared to deep learning as there is too little data in this dataset for this method.

Fig. 4 shows the detection results using precision-recall curves for the different methods. The best detection results are obtained with LOF-DTW, LOF-MSIG4 and SMK4. However, MSIG4 is 10 times faster than DTW (see Section IV-C) and completely unsupervised whereas SMK4 is semi-supervised. Note that MSIG4 provides better results than SIG4, indicating that the windowing operation is indeed useful for AD. Results obtained with OCSVM from [8] using DTW and SIG4 are also displayed showing a reduced detection performance.

TABLE I: Real datasets sizes and contamination rates.

| Dataset name | Nb time series | Nb points | Nb of anomalies |
|---|---|---|---|
| ANN Thyroid | 6916 | 21 | 250 (0.04%) |
| Breast Cancer | 367 | 30 | 10 (2.72%) |
| Letter | 1600 | 32 | 100 (6.25%) |
| Pen Global | 809 | 16 | 90 (11.12%) |
| Satellite | 5100 | 36 | 75 (1.47%) |

### B. Real data

This section analyzes data recovered from the unsupervised AD benchmark of the Harvard dataverse repository[1]. Dataset sizes and contamination rates are reported in Table I. This dataset has been used in previous AD papers such as [1] where a OCSVM with DTW baseline is provided. Table II shows the detection results of LOF using various similarity measures: Euclidean distance and DTW as baseline, and MSIG4. Overall, the MSIG provides promising AD results when compared to the baseline. Indeed, $\mathrm{F}_1$ scores are larger for three datasets out of five, and comparable to the best score on the remaining two. Thus, MSIG seems to be consistently efficient for detecting anomalies in these datasets. Note that contamination rates have to be taken into account, as for instance, there are only ten outliers to detect in the Breast Cancer dataset. Thus the recall score can only take ten different values, leading to large discrepancies in the computed $\mathrm{F}_1$ scores. The OCSVM method was also tested using the same similarity measures. Conclusions were similar to the ones obtained with LOF, but all detection scores were strictly inferior, similarly to the results of IV-A. An important aspect of the MSIG method is its low computational complexity, which is evaluated in what follows.

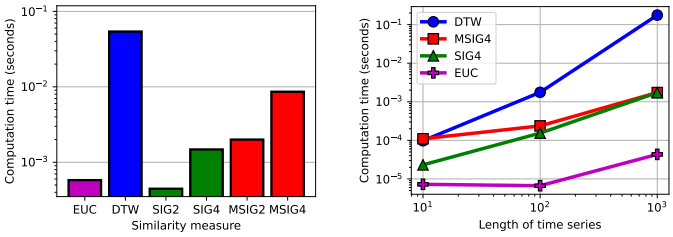
### C. Computation time and memory space

Given a $d$-dimensional time series $(X(t))_{t=1,\ldots,T}$ and a signature level $L \in \mathbb{N}$, the runtime complexity of the signature

[1] Available at https://doi.org/10.7910/DVN/OPQMVF.

TABLE II: $F_1$-scores using Local Outlier Factor (LOF) with 3 similarity measures and true contamination rates. Best in bold.

| | EUC | DTW | MSIG4 |
|---|---|---|---|
| ANN Thyroid | 0.09 | **0.22** | 0.17 |
| Breast Cancer | 0.60 | 0.60 | **0.70** |
| Letter | **0.55** | 0.44 | 0.52 |
| Pen Global | 0.58 | 0.43 | **0.59** |
| Satellite | **0.57** | 0.55 | **0.57** |




(a) Runtimes of pairwise distance matrix computations (synthetic data, 260 trajectories of size $2 \times 16$). SIG2 is the signature up to level 2. Log scale.

(b) Runtimes of pairwise distance matrix computation on randomly generated data. Each point is the average runtime over 1000 iterations. Log-log scale.

Fig. 5: Runtimes of the AD methods.

up to level $L$ (i.e., of $\mathbf{S}_{\leq L}(X)$) is $\mathcal{O}(Td^L)$ which has to be compared with the quadratic in length $\mathcal{O}(T^2d)$ runtime complexity of DTW. The space complexity of the signature is $\mathcal{O}(d^L)$, which does not depend on $T$. Therefore, for large values of $T$, the signature can be viewed as a compression tool. Fig. 5a displays runtimes for the computation of the pairwise similarity matrix. Note that the values shown for SIG/MSIG are the runtimes for the computation of the signature and the similarity matrix. The Euclidean distance (EUC) and SIG2 have complexities of the same order of magnitude. This is because the overhead of the computation of the signature is balanced with the Euclidean distance computation of a smaller number of signature coefficients. However, signature and multi-scale signature methods are at least one order of magnitude faster than DTW.

Fig. 5b compares runtimes on 10 randomly generated time series of dimension 2 for varying lengths $T \in \{10, 100, 1000\}$. As $T$ increases, the overhead created by the multi-scaling vanishes explaining why SIG4 and MSIG4 have the same complexity. This figure also shows that the proposed AD approach based on the multi-scale signature is numerically efficient, even for large datasets.

## V. Conclusions and Perspectives

This paper studied a new anomaly detection (AD) method for time series based on their multi-scale embedding in the space of signature features. A comparison between the proposed signature-based AD method and DTW on both simulated and real datasets showed similar detection performances. However, signature features can be computed multiple orders of magnitude faster than DTW, which is important for practical applications. The studied datasets had outliers that were globally abnormal. In the future, the proposed method should be applied to trajectories that contain collective anomalies (small abnormal segments), contextual anomalies or point anomalies [2]. A feature importance analysis could be used to detect which segmentation is most valuable to improve performances.

## References


[1] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tourneret, "How to introduce expert feedback in one-class support vector machines for anomaly detection?," *Signal Processing*, vol. 188, pp. 108197, 2021.

[2] B. Pilastre, L. Boussouf, S. D'Escrivan, and J.-Y. Tourneret, "Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning," *Signal Processing*, vol. 168, pp. 107320, 2020.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[4] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 6, pp. 1–39, 2012.

[5] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM Sigmod Record*, vol. 29, no. 2, pp. 93–104, 2000.

[6] 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*, vol. 13, no. 7, pp. 1443–1471, 2001.

[7] F. Meng, G. Yuan, S. Lv, Z. Wang, and S. Xia, "An overview on trajectory outlier detection," *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2437–2456, 2019.

[8] V. Mangé, J.-Y. Tourneret, F. Vincent, L. Mirambell, F. Manzoni Vieira, and B. Pilastre, "Détection de comportements anormaux dans des trajectoires de navires avec one-class SVM et dynamic time warping," in *Proc. Conf. GRETSI*, Grenoble, France, 2023, pp. 1–4.

[9] K.-T. Chen, "Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula," *Ann. Math.*, vol. 65, no. 1, pp. 163–178, 1957.

[10] T. Lyons, "Differential equations driven by rough signals," *Revista Matemática Iberoamericana*, vol. 14, pp. 215–310, 1998.

[11] T. Lyons and A. D. McLeod, "Signature methods in machine learning," *Preprint arXiv*, 2024.

[12] W. Yang, L. Jin, and M. Liu, "Chinese character-level writer identification using path signature feature, DropStroke and deep CNN," in *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, Tunis, Tunisia, 2015, pp. 546–550.

[13] J. Morrill, A. Kormilitzin, A. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons, "The signature-based model for early detection of sepsis from electronic health records in the intensive care unit," in *Proc. Int. Conf. Computing in Cardiology (CinC)*, Singapore, Singapore, 2019, pp. 1–4.

[14] N. Sugiura and S. Hosoda, "Machine learning technique using the signature method for automated quality control of Argo profiles," *Earth and Space Science*, vol. 7, no. 9, pp. e2019EA001019, 2020.

[15] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS One*, vol. 10, no. 3, pp. 1–21, 2015.

[16] I. Chevyrev and H. Oberhauser, "Signature moments to characterize laws of stochastic processes," *J. Mach. Learn. Res.*, vol. 23, no. 176, pp. 1–42, 2022.

[17] J. Diehl and J. Reizenstein, "Invariants of multidimensional time series based on their iterated-integral signature," *Acta Applicandae Mathematicae*, vol. 164, no. 1, pp. 83–122, 2019.

[18] B. Hambly and T. Lyons, "Uniqueness for the signature of a path of bounded variation and the reduced path group," *Ann. Math.*, vol. 171, no. 1, pp. 109–167, 2010.

[19] 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," *Preprint arXiv*, 2020.

[20] J. Morrill, A. Fermanian, P. Kidger, and T. Lyons, "A generalised signature method for multivariate time series feature extraction," *Preprint arXiv*, 2020.

[21] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005.

[22] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anomalous event detection," *IEEE Trans. Circuits Sys. Video Technol.*, vol. 18, no. 11, pp. 1544–1554, 2008.