Mechanical Systems and Signal Processing 224 (2025) 112048

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Real-time state estimation of nonstationary systems through dominant fundamental frequency using topological data analysis features

Arman Razmarashooli [a,*], Yang Kang Chua [b], Vahid Barzegar [a], Daniel Salazar [a], Simon Laflamme [a,c], Chao Hu [b], Austin R.J. Downey [d], Jacob Dodson [e], Paul T. Schrader [f]

[a] Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, USA
[b] Department of Mechanical Engineering, University of Connecticut, Storrs, CT, USA
[c] Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA
[d] Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA
[e] Air Force Research Laboratory, Munitions Directorate, Fuzes Branch, Eglin Air Force Base, FL, USA
[f] Air Force Research Laboratory, Information Directorate, Information Exploitation Branch, NY, USA

ARTICLE INFO

Communicated by L. Mevel

Keywords:
Structural health monitoring
High-rate systems
Nonstationary time series
Topological data analysis
Persistent homology
TDA features

ABSTRACT

Topological data analysis (TDA) is gaining popularity for classifying complex time series. Its integration with machine learning algorithm architectures shows promise in advancing predictive capabilities for various dynamic systems. Of interest to this paper is real-time applicability of TDA, targeting high-rate engineering systems. The study focuses on identifying a moving boundary condition on a fixed roller beam by analyzing the system's frequency. Point clouds are constructed from collected time series measurements based on the embedding theorem, and a time series of TDA features is created by employing a set of two sliding windows. From these features, the maximum persistence of homology groups is identified as one that is promising to perform real-time time series identification and is applied to laboratory datasets obtained from the DROPBEAR (Dynamic Reproduction of Projectile in Ballistic Environments for Advanced Research) testbed. Results demonstrate that the maximum persistence of the 0^{th} and 1^{st} dimensional persistence homology groups, namely H_0 and H_1, can provide a stable estimation of the cart location and have a lower noise level than other TDA features, with the maximum persistence of H_1 outperforming H_0. The maximum persistence of the 2^{nd} dimensional persistent homology group, H_2, is shown to be useful for detecting noise created by impact loads. Results also indicate that TDA features can be used to track the cart location within an acceptable range and perform similarly to or better than a short-time Fourier transform for more rapid dynamics, thus showing promise for real-time applications.

1. Introduction

Algebraic topology is a powerful method to extract pseudo-states capable of representing the essential dynamics of a system. The extraction of these pseudo-states is often based on Takens' embedding theorem [1]. Once extracted, the pseudo-states can be mapped

* Corresponding author.
E-mail address: shooli@iastate.edu (A. Razmarashooli).

https://doi.org/10.1016/j.ymssp.2024.112048
Received 11 January 2024; Received in revised form 11 September 2024; Accepted 14 October 2024
Available online 30 October 2024

to the dynamic states of the system of interest using a function that is frequently constructed using machine learning. Topological data analysis (TDA) combines algebraic topology with modern mathematical tools to further analyze the shape of data and is seen as a more robust method compared with pure algebraic topology-based strategies [2,3]. TDA has been successfully applied to analyze complex and chaotic dynamic systems [4]. Specifically leveraging persistent homology, TDA can be used to detect bifurcations within mechanical dynamics [5] and periodicity in time series [6–8].

Over the last two decades, TDA features have been integrated into machine learning algorithms such as neural networks to classify complex problems across various applications, including electrocardiograms [9,10], physiological data [11], structural health monitoring [12], and showed improvements in time series classification over conventional approaches [13–15].

Of interest to this paper is the real-time state estimation of dynamic systems using time series measurements, in particular for high-rate dynamic systems. High-rate dynamic systems are defined as systems experiencing dynamic events of amplitudes higher than 100 g_n, where g_n is the gravitational pull near the surface of the earth, over a duration of less than 100 ms (ms). Examples of high-rate systems include active blast mitigation systems, advanced weaponry, automotive airbag deployment mechanisms, and hypersonic vehicles [16,17]. A critical motivation in conducting real-time state estimation for these systems is in enabling feedback mechanisms that can be used to improve operational performance and safety. However, these systems typically exhibit (1) large uncertainties in the external loads, (2) high levels of nonstationary behavior and heavy disturbances, and (3) unmodeled dynamics generated from changes in the system configurations, which dramatically complicates the state estimation task. Existing efforts in the literature conducting real-time state estimation for high-rate dynamics include work from Joyce et al. [18] that developed a sliding mode observer-based algorithm with recursive least square parameters to accurately identify fundamental frequencies of the system, Downey et al. [19] that studied an algorithm to identify structural frequencies in real-time by minimizing errors within a set of finite elements, and where Wu and Todd [20] applied a damage identification framework using Gaussian process regression within a nonlinear autoregressive model, incorporating real-time weight updates. Additionally, Yan et al. [21] introduced an algorithm based on a model reference adaptive system capable of sub-millisecond computing capabilities.

All of these examples were based on physical representations. Here, we are concerned with data-driven techniques mapping time series measurements to structural states and propose leveraging TDA features in this endeavor. Yet, a notable challenge in applying topology-based techniques is in the restrictive nature of algebraic topology to stationary problems. We have proposed in [22] to forecast high-rate systems by decomposing a nonstationary system into a set of pseudo-stationary systems, and using an ensemble of recurrent neural networks (RNNs) to combine weighted predictions. This decomposition was achieved by creating a set of delay vectors, each of them comprising different essential dynamics of the system. The method showed great promise at forecasting nonstationary time series and was later combined with the model reference adaptive system-based algorithm proposed by [21] to create a state estimator using step-ahead predictions, thus effectively eliminating computing time lag [23].

In this paper, we investigate whether TDA features extracted from point clouds formed by delay vectors could be used as features to better explain the essential dynamics of a high-rate system, ideally providing a direct link from the point clouds to dynamic states. Hence, we study a technique to extract TDA features in real-time using two sliding windows to form a time series of TDA features. The technique is presented in the context of a single-harmonic-dominated system with changing boundary conditions, with applications to systems of higher dimensions left to future work. To do so, we first visit a set of TDA features that could be utilized with time series and select key TDA features to be studied. After, we investigate how TDA feature extraction can be performed sequentially and how point clouds from time series data can be effectively constructed to extract meaningful values.

To empirically support these objectives, numerical studies are performed on laboratory datasets obtained from a testbed termed dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR). The testbed consists of a cantilevered beam equipped with a fast rolling cart that is used to reproduce fast changes in boundary conditions. Our application problem of interest is the real-time estimation of the moving boundary condition on the cantilever beam through the movement of the cart. We chose this problem because (1) experimental datasets have been made publicly available [24], and (2) DROPBEAR datasets are well suited for validating and benchmarking performance of high-rate algorithms [24]. Note that the authors' intent is to eventually apply TDA in the high-rate realm, thus providing a time-based data analysis method that could surpass frequency-based methods in terms of computing time, also anticipating that the use of TDA features could significantly ameliorate the convergence of on-the-edge machine learning representations. However, the issue of computing time, which, when conducting state-of-the-art TDA computations, does not fall within the time frames necessary for high-rate systems health monitoring, is ongoing and therefore left for future work.

The rest of the paper is organized as follows. Section 2 provides background information on the application of TDA. Section 3 describes the algorithm developed for deploying TDA in real-time, as well as the DROPBEAR testbed. Section 4 presents and discusses the results obtained from synthetic and experimental datasets. Section 5 reports conclusions and recommendations.

2. Background

This section presents the background of TDA, how it can be applied to time series data, and defines key TDA features of interest.

TDA can be used to study the topology of point clouds using fundamental concepts of modern mathematics [25] which exploit the categorical (or functorial) relationship between topology and algebra. This is often done using homology theory, where homology groups and group homomorphisms are induced by associated topological spaces and the continuous mappings between them. For example, Euclidean space $\mathbb{R}^n$ and its subspaces, $\mathbb{R}^{n-i}$ where $n, i \in \mathbb{Z}_{>0}$ and $i < n$ have a topological structure induced by the Euclidean metric and subspace topologies. The inclusion maps $i : \mathbb{R}^{n-i} \hookrightarrow \mathbb{R}^n$ between those spaces are continuous. This induces a matching homological group representation of finite dimensional (quotient) vector spaces and linear transformations. The homology groups

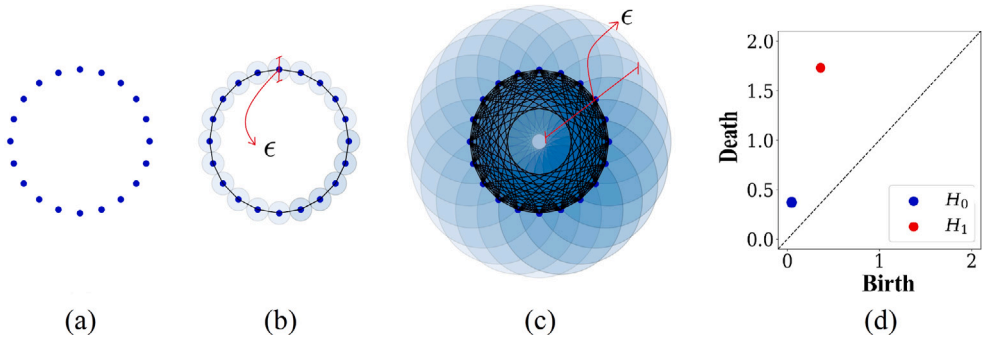

Fig. 1. The filtration of Rips complex. There are 20 points and 20 individual connected components in (a) (feature H_0). As the balls of radius ϵ expand, they connect and merge into one connected component. Eventually, all of the points connect; thus, H_0 remains one connected component for all time, and a hole (feature H_1) is initially born within the connected components for $\epsilon = 0.32$ (b). This hole dies for $\epsilon \geq 1.79$ as shown in (c). The associated persistence diagram (PD) is plotted in (d).

track, measure, and algebraically encode a given data's topological characteristics such as connected components, holes, and voids, allowing for ingestion into algorithmic/automated implementations. These groups are denoted as $H_k(X)$, where X is a topological space and k is the homological dimension. For instance, $k = 0$ algebraically represents and encodes the connected components of X, $k = 1$ its 1-dimensional holes, $k = 2$ its 2-dimensional holes or voids, etc.

Similarly in computational topology, one characterizes the topological space for a finite point cloud (FPC) embedded in a metric space such as $\mathbb{R}^n$ analyzing characteristics of its induced homology groups [26]. For example, one can determine the order or rank of its linearly independent generators in each homological dimension. In effect, this "counts" the number of connected components of the topological space X in $H_0(X)$, the number of 1-dimensional holes of X in $H_1(X)$, the number of voids of X in $H_2(X)$, etc. These counts, which are also known to be topological invariants, are often called the kth-Betti numbers for each $H_k(X)$. Generation of the underlying spaces associated to these homology groups can be done through a variety of methods, but often starts with the construction of a simplicial complex filtration that creates an evolution of inclusive subspaces using simplices whose vertices are elements of the FPC [25]. In this manner a filtration assigned to the FPC can be utilized to define its topological invariants. A filtration's simplicial complexes associated with the FPC are dependent on the $\mathbb{R}$-valued scale parameter ϵ, which is the evaluation of the embedding space's metric between the FPC's points. For example, in $\mathbb{R}^n$, ϵ would be the evaluation of the n-dimensional Euclidean distance between two given points of the embedded FPC. There are two major methods to associate a simplicial complex with its underlying point cloud data, namely the Cech and Vietoris–Rips complexes [25]. In this paper, our method utilizes the Vietoris–Rips complex due to its computation efficiency for large datasets. With the Vietoris–Rips method, a group of points spans a k-simplex if and only if pairwise distance between FCP vertices is less than or equal to ϵ [25]. The interested reader will find a more formal mathematical definition of TDA in the Appendix, including the use of Vietoris–Rips complex as a filtration technique.

In practical applications, determining the value of ϵ to identify topological features of interest in an FPC can be elusive or intractable [25]. To address this challenge, one can use the above-mentioned filtration methodology, which continuously dilates the $\mathbb{R}$-valued $\epsilon > 0$, and for each ϵ, generates simplicial complexes. As we let $\epsilon \to \infty$, new topological features represented in the complexes based on the original manifold may appear or disappear, and the homology groups encode those changes for further processing. Persistent homology (PH) tracks these changes by associating each topological feature with a birth and death time as $\mathbb{R}_{\geq 0}$-valued order pairs or $(birth, death) = (b, d)$ pairs. The total output is a persistent homology dimension based multi-set of all topological features represented by (b, d) pairs, where $b < d$, can be visualized in the form of a persistence diagram (PD) [27]. These diagrams can be used to extract several PD based comparative metrics or representations, including bottleneck and Wasserstein distances metrics, the persistence landscape and silhouette representations, and the number of off-diagonal points metric [25]. To illustrate, Fig. 1 shows the filtration of Vietoris–Rips complexes obtained from a point cloud consisting of 20 equally scattered points in space, each surrounded by expanding balls of radius ϵ. Initially, there are 20 connected components, one for each point, forming the initial 0th persistent homology dimension H_0 (Fig. 1(a)). As the ϵ-radius of the balls increases, they create a simplicial complex, here for $\epsilon = 0.32$. For that radius, because all points are equidistant, H_0 represents a single connected component, and H_1 begins to track and measure the lifespan from the creation of a hole (Fig. 1(b)). This hole H_1 then dies at $\epsilon = 1.79$ (Fig. 1(c)). The associated PD that tracks the evolution of H_0 and H_1 is plotted in Fig. 1(d).

2.1. TDA features of interest

Several TDA based features can be extracted from the persistence diagram. Of interest to this paper are those that can be applied to the study of time series. They are briefly described in what follows. In this work, we make use of the Giotto-TDA Python package, which offers a more straightforward application programming interface [28].

Persistence Diagram: Let D_1 and D_2 be two persistence diagrams where each diagram consists of a set of points in the plane. The points in D_1 are denoted as $\{x_i(b_i, d_i)\}$ for $i = 1, 2, \ldots, n$, where x_i represents a point, b_i represents the birth time (birth of a

topological feature) and d_i represents the death time (death of a topological feature) of the ith feature. Similarly, the points in D_2 are denoted as $\{x_j(b_j, d_j)\}$ for $j = 1, 2, \ldots, m$. A trivial diagram, denoted as D_t, is an empty diagram that has no points. In other words, D_t contains no features and is represented as an empty set $D_t = \emptyset$.

Maximum Persistence: Maximum persistence is the longest persistence time (death−birth) for a given topological feature (connected components, loops, voids, etc.) over a range of scale parameters. Maximum persistence features are considered the most robust and meaningful features, as they are less likely to be artifacts of noise or small-scale fluctuations in the data. This feature can identify significant and stable topological features in the data [29].

Bottleneck Distance: Bottleneck distance is a metric used to compare two persistence diagrams and is defined as the infimum of the maximum distance between matched points in the two diagrams. Cohen-Steiner et al. introduced the Bottleneck distance and proved its stability, whereas small perturbation of the input data lead to small changes in the persistence diagram [30]. When calculating the bottleneck distance, one must find the largest distance (L_∞ norm) between corresponding points in two persistence diagrams by considering all possible pairings [3]. The bottleneck distance between two persistence diagrams D_1 and D_2, denoted as $d_B(D_1, D_2)$, is defined as:

$$d_B(D_1, D_2) = \inf_{\gamma} \max_{x \in D_1} \{\|x - \gamma(x)\|_\infty\} \tag{1}$$

Here $\|x - \gamma(x)\|_\infty$ represents the supremum norm (L_∞ norm) of the distance between the points in the persistence diagram D_1 and their corresponding matched points in D_2 under the bijection $\gamma : D_1 \to D_2$. The infimum is taken over all possible bijections [31].

To evaluate our system's effectiveness based on a single diagram, we must define the Bottleneck Amplitude, which is calculated by measuring the distance between two diagrams, one of which is trivial. Thus, in this scenario, the perfect matching between two diagrams is obvious and we can relate it to the maximum persistence.

$$d_B(D_1, D_t) = \frac{1}{\sqrt{2}} \max_{x \in D_1} (d_i - b_i) \tag{2}$$

Wasserstein Distance: The Wasserstein distance is a distance metric that measures the difference between two persistence diagrams, D_1 and D_2. Unlike the Bottleneck distance, which only considers the most significant feature and is a worst-case scenario metric, the Wasserstein distance takes into account all features in the diagrams. This makes it more sensitive to differences across all features [32]. To calculate the Wasserstein distance, we first find the pth power of the distance between corresponding points in the diagrams, then take the pth root of the sum of those distances.

$$W_p(D_1, D_2) = \left(\inf_{\gamma} \sum_{x \in D_1} \|x - \gamma(x)\|_\infty^p \right)^{1/p} \tag{3}$$

where γ ranges over all possible bijections between the points of D_1 and D_2, and p is a real positive number, typically chosen 1 (earth mover's distance) or 2 (Euclidean distance).

Similar to the Bottleneck amplitude, we can define the Wasserstein Amplitude, representing the Wasserstein distance from a trivial diagram for $p = 1$ [3].

$$W_1(D_1, D_t) = \frac{1}{\sqrt{2}} \sum_{x \in D_1} (d_i - b_i) \tag{4}$$

Persistence Landscape: The persistence landscape is a metric used in TDA to summarize and compare persistence diagrams. Persistence landscapes map persistence diagrams into a function space, which can then be studied using tools from statistics and machine learning [33]. The persistence landscape for diagram D_1 is a sequence of functions $\lambda_k : \mathbb{R} \to [0, \infty]$, with k ranging from one to infinity, where each function $\lambda_k(x)$ is defined as the kth largest value of the piecewise linear function $f(b_i, d_i)(y)$ [33]:

$$f(b_i, d_i)(y) = \begin{cases} 0 & \text{if } y \notin (b_i, d_i) \\ y - b_i & \text{if } y \in [b_i, (b_i + d_i)/2] \\ -y + d_i & \text{if } y \in [(b_i + d_i)/2, d_i] \end{cases} \tag{5}$$

with the persistence landscape taken as:

$$\lambda_k(y) = \max_i f(b_i, d_i)(y) \tag{6}$$

The amplitude calculation for persistence landscapes involves computing the L^2 norm distance from the persistence landscape λ to the corresponding closest diagonal point in the trivial persistence diagram λ' [31]. This distance is given by:

$$\|\lambda - \lambda'\|_2 = \left(\sum_{k=1}^{N} \int |\lambda_k(y) - \lambda'_k(y)|^2 \, dy \right)^{1/2} \tag{7}$$

where N is the number of layers in the persistence landscape.

Persistence Silhouette: The persistence silhouette is a metric that quantifies the separation and quality of persistent homology classes in a dataset. It provides a measure of how well-defined and distinct the topological features are in a persistence diagram. This concept is closely related to the persistence landscape, which characterizes the evolution of these features across different scales [34].

The persistence landscape ($\lambda_j(x)$) represents the persistence of features as functions of a parameter (often associated with scale). To capture the separation of features at different scales, we introduce a weighted average of the landscape functions:

$$\Phi(x) = \frac{\sum_{j=1}^{n} \omega_j \lambda_j(x)}{\sum_{j=1}^{n} \omega_j} \tag{8}$$

To prioritize the importance of features with greater persistence values, we set the weights (ω_j) as $\omega_j = |d_j - b_j|^p$ for $0 < p < \infty$. This leads to the definition of the power-weighted silhouette:

$$\Phi^{(P)} = \frac{\sum_{j=1}^{n} |d_j - b_j|^p \lambda_j(x)}{\sum_{j=1}^{n} |d_j - b_j|^p} \tag{9}$$

The amplitude of the silhouette can be calculated by measuring the L^2 norm distance between λ and λ', representing the silhouette and the trivial silhouette, respectively, at each uniform sample:

$$\|\lambda - \lambda'\|_2 = \left(\sum_{k=1}^{N} \int |\lambda_k(x) - \lambda'_k(x)|^2 \, dx \right)^{1/2} \tag{10}$$

Here, N is the number of layers in the persistence landscape.

Number of Off-Diagonal Points: This feature calculates the number of points above the diagonal line that persist longer than zero over each homology dimension in the persistence diagram [3]. Counting the number of off-diagonal points can be a useful way to assess the complexity of the dataset and the presence of persistent topological features that are not immediately short-lived. For example, the number of off-diagonal points for H_0 reflects the sampling rate of the point cloud and corresponds to the number of points in the point cloud, and the number of off-diagonal points for H_1 and H_2 indicates the number of holes and voids present in the point cloud.

3. Real-time TDA feature extraction

In this section, the proposed algorithm for real-time applications is presented and applied to the DROPBEAR datasets. First, the concept of transforming time series into point clouds and multi-resolution windowing for nonstationary signals is presented. After, the DROPBEAR testbed is presented along with its simplified model along with a physical approach to the interpretation of the TDA features to provide the context for the construction of the real-time feature extraction algorithm.

3.1. Transforming time series into point clouds

Experimental and field conditions usually involve the measurement of one or many states using strategically placed sensors, and these measurements are made available in the form of time series. In order to conduct TDA, these time series need to be transformed into an embedded point cloud, analogous to the FPC discussed in the prior section. A well accepted technique to do so is based on Takens' embedding theorem [1] and its extensions to non-autonomous systems with deterministic forcing [35], state-dependent forcing [36], stochastic forcing [37], and more recently to multivariate datasets [38]. It consists of embedding a given measurement vector $\mathbf{x}(t)$ into a delay vector $\chi(t) = [\mathbf{x}(t),\ \mathbf{x}(t-\tau),\ \mathbf{x}(t-2\tau),\ \ldots,\ \mathbf{x}(t-(\mu-1)\tau)]$ of embedding dimension μ using measurements delayed by time τ. The embedding theorem states that given the appropriate selection of μ and τ, χ will generate an embedding topologically equivalent to the original dynamic system. Hence, χ can be utilized to compute topological invariants of the original system. Here, a time series is transformed into a point cloud X by assembling the delay vectors $\mathbf{X}(t) = [\chi(t),\ \chi(t-\tau),\ \chi(t-2\tau),\ \ldots,\ \chi(t-(k-1)\tau)]$ with k consistent with μ and the size of $\mathbf{x}(t)$. A critical constraint is that the embedding theorem only applies to stationary systems, and our dynamics of interest is nonstationary. As a solution, we use a sliding window w restricted to the measurements of the dataset $\mathbf{x}_w(t) = [x(t),\ x(t-1),\ x(t-2),\ \ldots,\ x(t-w)]$ and assume that the dynamics are stationary within w [31]. This assumption is based on the knowledge that the dynamic system returns to a stationary state after disturbance. Thus, the point cloud only includes data within w, where k is now consistent with μ and the size of $\mathbf{x}_w(t)$.

With this technique, TDA is based on the choices of μ, τ, and w. In theory, the embedding dimension is selected based on the attractor size n, with $\mu > 2n$ which is not a necessary condition. In practice, n is often unknown, and it is common to over-embed the delay vector. Time delay τ needs to be selected such that the phase space formed by the point cloud unfolds sufficiently to contain topological information, but before it starts to fold back into itself where topological information starts to get lost. Typically, μ is selected using the false nearest neighbor test, and τ through mutual information. The utilization of a sliding window w is not typical, because the vast majority of work utilizing the embedding theorem is based on stationary systems.

3.2. Multi-resolution windowing for nonstationary signals

Embedding of whole time series is practically impossible due to significant computational load needed for creating simplicial complexes. Using the windowing method discussed above, one can extract TDA features in real-time, with a lag that is proportional to the window size. For a single harmonic signal, the optimal time delay that provides data spread is $\tau = 0.25/f$, where f is the given signal's frequency [12,39]. When reconstructing the phase space of harmonic signals using the optimum time delay $\tau = 0.25/f$, the resulting shape is consistently a unit circle.

Literature has shown that the ideal embedding dimension for a single harmonic signal is $\mu = 2$ [8]. Nonetheless, to account for noise, results have shown that a dimension $\mu = 3$ was more appropriate. A 3-dimensional point cloud emerges as a result, along with its related persistence features of connected components or zero-dimensional holes (TDA feature H_0), loops or one-dimensional holes (TDA feature H_1), and voids or two-dimensional holes (TDA feature H_2). When constructing persistence diagrams using simulated, noise-free data, feature H_1 is born when feature H_0 dies and feature H_2 is born when feature H_1 dies. Note that the death time and the number of points related to the H_0 feature depend on the number of points in the point clouds and their persistence interval length converges to zero as the sampling rate increases. The maximum persistence of H_0 represents the greatest spatial separation between points within the point cloud. While the maximum distance remains constant for a fixed frequency and time delay, the distances between points will change with a varying frequency f, subsequently impacting the maximum persistence of H_0. This variability, when maintaining a fixed sampling rate, can be interpreted as a representation of the frequency of the system, as the points become more dispersed within the point cloud. In systems characterized by high sampling rates, the length of all persistence intervals of H_0 tends to approach zero. Even as this value decreases, the normalized value of the maximum persistence of H_0 remains a meaningful indicator of the system's frequency, although the maximum persistence of H_0 is very sensitive to noise and not the best choice to the interpreting systems with a high range of frequency and noise.

The existence of H_1 represents the periodicity in signal and characterizes the evolution of 2D holes within the point cloud. The maximum persistence of H_1 for the unit circle is either equal to or greater than $\sqrt{3}$, as noted in [40]. However, when the time delay falls below the optimal value ($\tau < 0.25/f$), the reconstructed phase space takes the form of an ellipse. The ratio of the ellipse's major axis to its minor axis becomes crucial in determining the maximum persistence of H_1. A more circular ellipse corresponds to a higher maximum persistence of H_1, as discussed in [41,42] and as the system's frequency increases, the hole within the ellipse formed by the point cloud persists longer, and so does the maximum persistence of H_1. In a noise-free scenario, the persistence diagram typically features a single point for H_1 representing the most significant hole. The birth time of H_1 depends on the sampling rate of the system since it defines the distance between points. As the sampling rate increases, meaning more data points are collected, the birth time of H_1 tends to zero. This study introduces a novel definition for maximum persistence, specifically accounting for scenarios where the birth time is zero. This adjustment aims to provide a clearer understanding of the persistence of topological features, mitigating the impact of sampling rate variations and point-to-point distances.

Regarding H_2, it is important to note that the feature in a noise-free scenario typically is absent, mainly due to the repetitive nature of the periodic function. In effect, the maximum persistence of H_2 can be used to describe the noise in the system.

In a nonstationary system, because the frequency f is expected to vary, we set a maximum allowable time delay $\tau_{\max} = 0.25/f_{\max}$ with $f_{\max}$ being the maximum frequency of the system. Using a higher τ would risk folding the topological space onto itself, leading to information loss. It is also important to select τ to allow the phase space to sufficiently unfold so as to generate topological features without being unreasonably small.

To analyze the time series, we employ two sliding windows with different sizes. The first window has a size of $w = 1/f_{\min} + 2\tau$, with $f_{\min}$ being the minimum frequency of the system. This ensures that the point cloud will form a complete loop, representing H_1. However, as the frequency of the system increases, loops are expected to overlap, thus affecting the value of the maximum persistence of H_0. To address this issue, the second window is defined to be smaller ($w = 1/f_{\max} + 2\tau$) for which points will not overlap. This approach guarantees that the window size is sufficient to capture the characteristics of all frequencies in the data.

3.3. Application to DROPBEAR

The DROPBEAR testbed was developed to validate state estimation algorithms for high-rate dynamic systems. The testbed (Fig. 2) is fully described in [18]. Briefly, it consists of a 505 mm cantilever (clamped) beam that is subjected to a fast-moving boundary condition, which is achieved using a movable cart. The cart is used to mimic sudden or gradual changes in stiffness. One accelerometer is installed on the beam located 300 mm away from the beam's fixed end to measure the response. The dynamics of the beam can be excited either by the moving cart alone or by applying an impact load using an impact hammer. Additionally, a mass is attached to the end of the beam using an electromagnet to explore how a change in mass affects systems dynamics. However, in our study, the mass remains attached. Past research efforts using DROPBEAR data have simplified the testbed as a cantilever beam with a time-varying boundary condition that varies its effective length L. In the context of free-vibration, the fundamental frequency ω_j of a beam can be estimated using Eq. (11) [43]:

$$\omega_j = 4\pi^2 \sqrt{\frac{EI}{\rho A}} \left(\frac{4j+1}{4L} \right)^2 \tag{11}$$

where ω_j is the natural frequency of the jth mode of vibration in rad/s, E is Young's modulus, I is the moment of inertia, ρ is density, A is the cross-sectional area, and L is the length of the beam. In our case, this equation can be used to map the fundamental frequency of the beam to its length. Under free vibration and assuming that the first mode dominates the response, the response of the beam can be simplified to be a single harmonic time series:

$$x(t) = A\cos(\omega t) = \cos(2\pi f t) \tag{12}$$

where A is the amplitude, $x(t)$ is the motion, f is the natural (first) cyclic frequency in Hz, and t is time in seconds.

The specialized case of DROPBEAR allows us to take a physical approach to the interpretation of the TDA features discussed in the previous section. Thus, under a fixed time delay τ, the frequency of the harmonic signal influences the spatial distance between points and the ratio of the major axis to the minor axis of the ellipse, subsequently affecting the maximum persistence of H_0 and

H_1. This observation implies that the maximum persistence of H_0 and H_1 can be associated with the frequency of a harmonic signal when a specific value of τ is chosen. In our case, when the system is disturbed by an impact caused by an impact hammer, H_2 should appear in the persistence diagram. When dealing with a simple harmonic signal, we can easily correlate most of the previously defined TDA features to the maximum persistence values within each homology group. As mentioned before, with a large enough τ and in the absence of noise with n points in the point cloud, off-diagonal points in the persistence diagram should consist of n H_0 points (n connected components), one H_1 point (a long persistence hole), and no H_2 point (no voids). The value of the Bottleneck Amplitude is determined by measuring the maximum distance to the diagonal line, which is equal to the maximum persistences of H_1 and H_0. The Wasserstein Amplitude along with Persistence Landscape and Silhouette are also influenced by the maximum persistence point's value and the number of off-diagonal points in the persistence diagram. Since we have only one point in the persistence diagram for H_1 over a constant frequency, their normalized value will be equal to the normalized maximum persistence. However there are some fluctuations in the persistence landscape and silhouette since the point cloud during a frequency transition will not form a full ellipse, but these features should still be similar to the maximum persistences. Thus, the maximum persistence of H_1 is expected to play a major role in shaping TDA features.

4. Numerical simulations

We study the performance of our windowing method to extract the physically meaningful TDA features related to the persistence interval of H_0 and H_1 for a nonstationary harmonic excitation. This study is conducted over two types of datasets: (1) a synthetic noise-free harmonic signal; and (2) experimental data from DROPBEAR.

4.1. Synthetic harmonic signal

The synthetic harmonic signal is taken as

$$x(t) = \cos\left(2\pi f(t)t\right) \tag{13}$$

with $f(t)$ varying between 20 and 50 Hz. The excitation is plotted in Figs. 3 and 4 (gray solid line). In this excitation, the frequency remains constant at 20 Hz for the first two seconds (0–2 s) before increasing to 50 Hz over the next two seconds (2–4 s), after which it remains at 50 Hz for another 2 s, and then decreases to 20 Hz and remains constant for another 2 s. The size of the moving windows is $w_1 = 1/f_{\min} + 2\tau = 0.05 + 0.002 = 0.052$ s for plotting the maximum persistence of H_1, and $w_2 = 1/f_{\max} + 2\tau = 0.02 + 0.002 = 0.022$ s for plotting the maximum persistence of H_0. Data are embedded using $\tau = 0.001$ s for both windows ($0.25/f_{\max} = 0.005$ s).

Figs. 3 and 4 plot the evolution of various TDA metrics related to the persistence interval of H_0 and H_1, respectively, compared against the frequency of the signal (solid blue line). For TDA features related to the persistence interval of H_0 (Fig. 3), we can observe that all features remain constant when the frequency is constant. Conversely, while the frequency is varying, all features exhibit noise, yet with a trend following the shift in frequency. Notably, there is a clear link between Wasserstein and Bottleneck amplitudes and maximum persistence being similar even in the transition part. During the shift from 20 Hz to 50 Hz, deviations in the point cloud shape lead to fluctuations in the Persistence Landscape and Silhouette metrics. Another key characteristic in the figure is the time lag in the features with respect to frequency which is equal to the window size. This lag is attributed to the sliding window method that takes measurements in the past, and diminishes when the frequency increases due to the more rapid creation of complete loops.

For TDA features related to the persistence interval of H_1 (Fig. 4), the plot exhibits more robust performance for all features, attributed to the presence of one long-lasting hole in the dataset that persists during frequency transitions. Here, the silhouette and landscape metrics tend to deviate slightly from the maximum persistence. The maximum persistence, along with the Wasserstein and Bottleneck amplitudes, relate more closely to the frequency. There is also a time lag in the features similar to that found over H_0 features. From these results, we can conclude that the maximum persistence of H_0 and H_1 can be useful in tracking a single-frequency system. These features, along with the maximum persistence of H_2, are selected in moving forward with DROPBEAR data.

4.2. DROPBEAR

The investigation is pursued using realistic datasets taken from the DROPBEAR experimental testbed. First, the evolution of TDA features is examined on two datasets with a dynamics resembling that of the synthetic dataset discussed in the previous section. Second, the study is repeated on a dataset with a more complex dynamics. Datasets are taken from the dataset bank available online [24], and a low pass filter at 150 Hz is employed prior to running the code in Python on an Intel(R) Xeon(R) CPU E3-1240 v5 3.50 GHz.

The first two datasets correspond to Dataset-6, where one experiment does not involve the use of an impact hammer (test 1) and the other does (test 2). In those tests, the cart is initially located 50 mm from the clamp, moves to 200 mm from the clamp, and comes back to its original position. The temporal location of the cart is plotted in Figs. 6 to 10 (blue line). The system's dominating frequency varies between 17.7 Hz (at 50 mm) and 31.0 Hz (at 200 mm). Accelerometer data was acquired at 25 kHz. Data is embedded using $\tau = 0.004$ s ($0.25/f_{\max} = 0.008$ s) and the size of the moving windows are taken as $1/f_{\min} + 2\tau = 0.06$ and $1/f_{\max} + 2\tau = 0.04$ s. The window slides every 0.001 s to lessen computational demands. Figs. 5 and 6 plot the maximum persistence for different homology groups for tests 1 and 2, respectively. Results are compared against those obtained using a short-time Fourier transform (STFT) conducted under a window size of 4096 samples (equivalent to 0.164 s) and a window length overlap of 0.008 s.

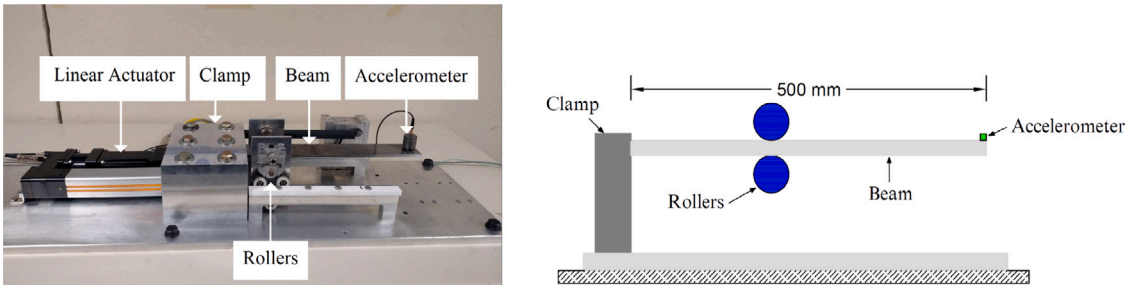

Fig. 2. Picture of the DROPBEAR testbed.

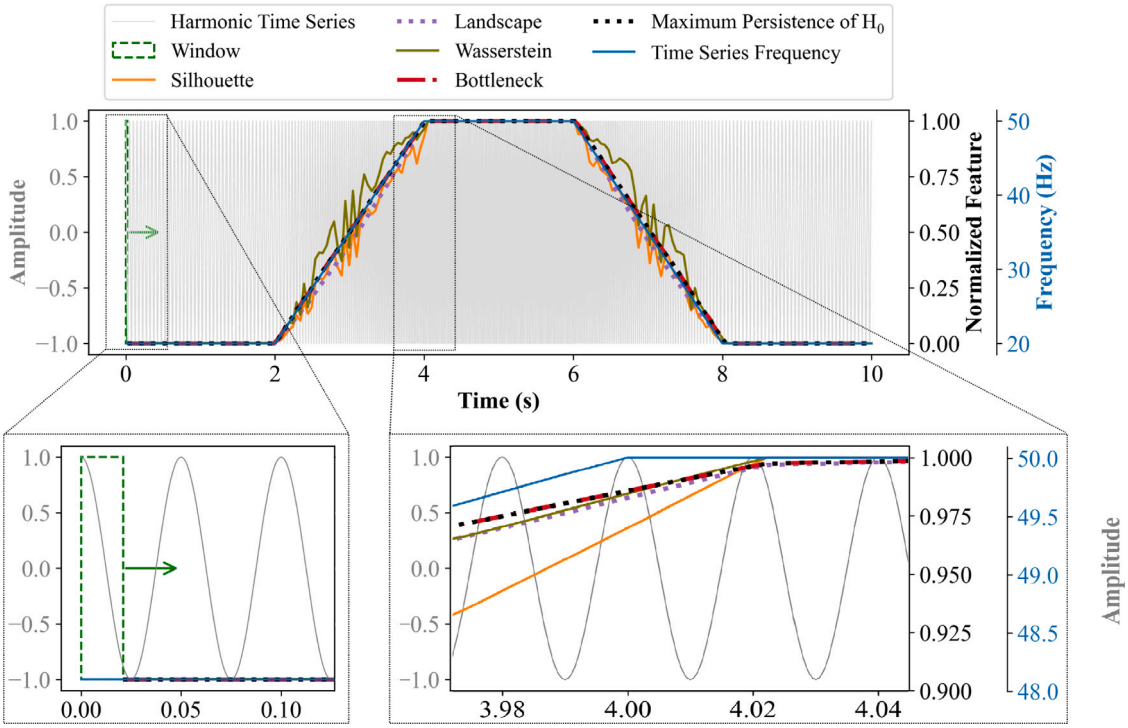

Fig. 3. Evolution of TDA features constructed from H_0 over a synthetic, noise-free harmonic signal.

Data obtained from the maximum persistence of H_0 and H_1 is post-processed to map results to the cart location. Combining the assumption that these TDA features relate linearly to frequency and the frequency-cart location relationship from Eq. (11), we conduct the following linear regression

$$x^2 = a_0 + a_1 H_i, \tag{14}$$

where x is the cart location in mm, a_0 and a_1 are regression parameters estimated using a least squares estimator (LSE), and H_i is the homology group for $i = 0, 1$. A similar process is conducted on STFT data to allow a better comparison. Data from the maximum persistence of H_2, because it does not relate to the system's frequency, is normalized between 0 (minimum value) and 1 (maximum value). In linear regression analysis of each dataset, the initial part of the excitation is removed because of the presence of high noise. For test 1, test 2, and test 3, the initial noisy parts are the first 0.74 s, 0.43 s, and 1.4 s of the excitation, respectively.

In test 1 (no impact hammer, Fig. 5), the initial part of the excitation returns very noisy TDA features, attributable to the low acceleration signal as the cart is yet to move. As the cart starts moving, both the maximum persistence of H_0 (shown by the purple dotted line) and H_1 (shown by the red dotted line) correlate with the cart location. At the end of the excitation, the cart has a slightly richer acceleration signal due to free vibrations after the last movement of the cart, thus returning TDA features exhibiting less noise than at the beginning. The maximum persistence of H_2 remains close to zero for most of the signal, except at the beginning during the low acceleration signal attributable to higher noise.

In test 2 (impact hammer, Fig. 6), all of the three maximum persistences exhibit less noise after the first hammer impact, attributable to the richer signal. Another difference with respect to test 1 is that the maximum persistence of H_1 is slightly more

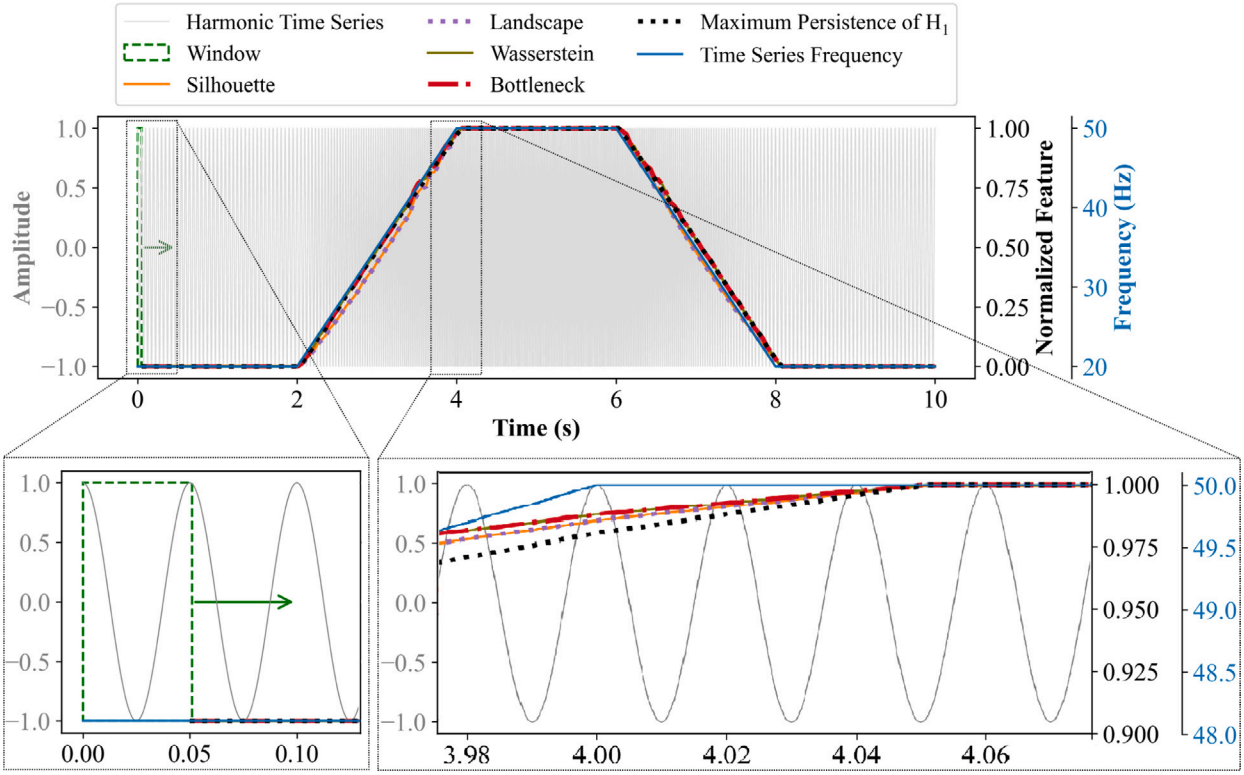

Fig. 4. Evolution of TDA features constructed from H_1 over a synthetic, noise-free harmonic signal.

noisy than that of H_0. Also, the maximum persistence of H_2 peaks during impacts attributable to the local signal being non-harmonic and can thus be used to identify the impacts themselves.

The third dataset (test 3) corresponds to Dataset-2 [44]. In this test, the cart movement is more complex and divided into three parts, and the experiment does not utilize the impact hammer. The cart is initially located 50 mm from the clamp, and moves back and forth to 70, 90, 110, 130, 150, and 170 mm from the clamp with different ramping and holding patterns, as shown in Fig. 7(blue line). The system's dominating frequency varies between 44 Hz (at 50 mm) and 96 Hz (at 170 mm). Accelerometer data was acquired at 12.5 kHz. Data is embedded using $\tau = 0.0008$ s ($0.25/f_{\max} = 0.002$ s) and the size of the moving windows are taken as $1/f_{\min} + 2\tau = 0.025$ s and $1/f_{\max} + 2\tau = 0.018$ s. The window slides every 0.001 s to lessen computational demands.

The difference between tests 1 and 2, and test 3 lies in the velocity and the acceleration of the cart. Both tests 1 and 2 have a maximum cart velocity of 120 mm/s and a maximum acceleration of 250 mm/s^2. In the second section of test 3, the maximum cart velocity and acceleration are more similar to tests 1 and 2, measuring 70 mm/s and 130 mm/s^2, respectively. However, in sections 1 and 3 of test 3, the maximum cart velocity and acceleration reach 250 mm/s (maximum allowable velocity of the test) and 3600 mm/s^2, thus exhibiting features closer to those of a high-rate system.

The evolution of TDA features for test 3 is plotted in Fig. 7, with Figs. 8 to 10 respectively zooming over the first, second, and third sections of the cart displacements. A notable observation is that, despite the lack of impacts, the maximum persistence of H_2 is often non-zero when the cart initiates movement, a behavior more noticeable over the first section of the displacements (Fig. 8), attributable to a higher initial acceleration of the cart compared to that in tests 1 and 2. Results also show that the maximum persistence of H_1 outperforms H_0.

Results from DROPBEAR show that the maximum persistence of H_0 and H_1 can be mapped to the cart location. Note that parameterizing the map is out of the scope of this research because we focus on studying whether TDA features can be used to explain the dynamics of a high-rate system. However, a rapid evaluation of the results of the regression conducted on Eq. (14) listed in Table 1 indicates that (1) the regression parameters remain within the same magnitude across test types, and (2) feature H_1 provides a better estimation of the cart location relative to H_0 when comparing the quality-of-fits (R^2).

Next, we further examine the performance of the maximum persistence of H_1 at mapping to the correct cart location. The mapping performance is assessed using three performance metrics (J_1, J_2, and J_3) over the truncated part of the signal, after removing the initial noisy segment. These metrics are constructed to denote improvement over decreasing numbers.

Metric J_1 is the mean absolute error between the actual cart location and the mapped cart location for the entire excitation or section of the excitation. Metric $J_{2,i}$ is the number of instances over the signal that the estimation does not remain below a given distance i (in mm, here taken as $i = 5, 10,$ and 20 mm, with 20 mm arbitrarily taken as an acceptable target). Metric $J_{3,i}$ is the mean

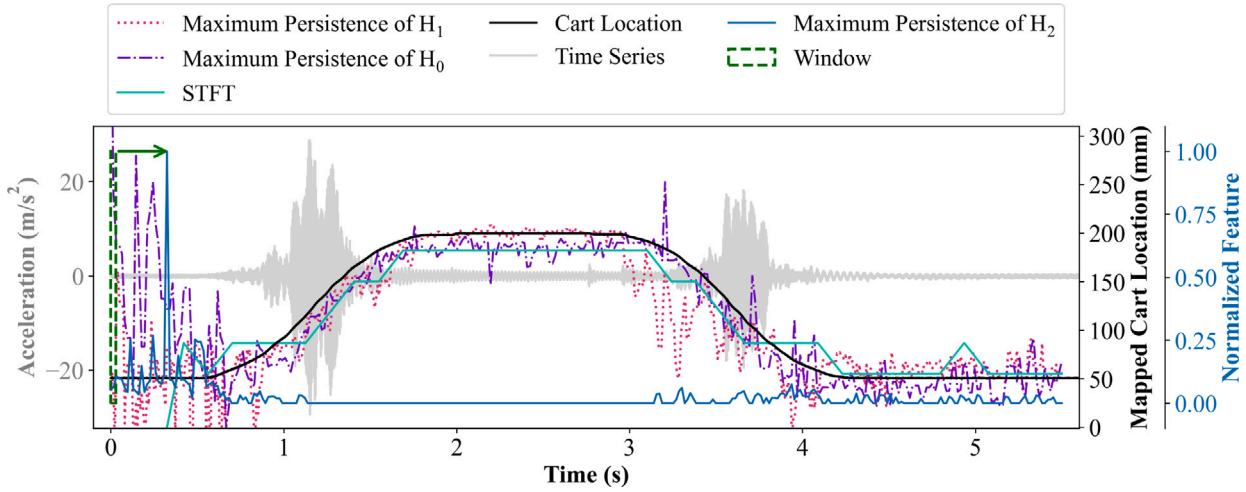

Fig. 5. Evolution of TDA features over test 1 (without impact hammer).

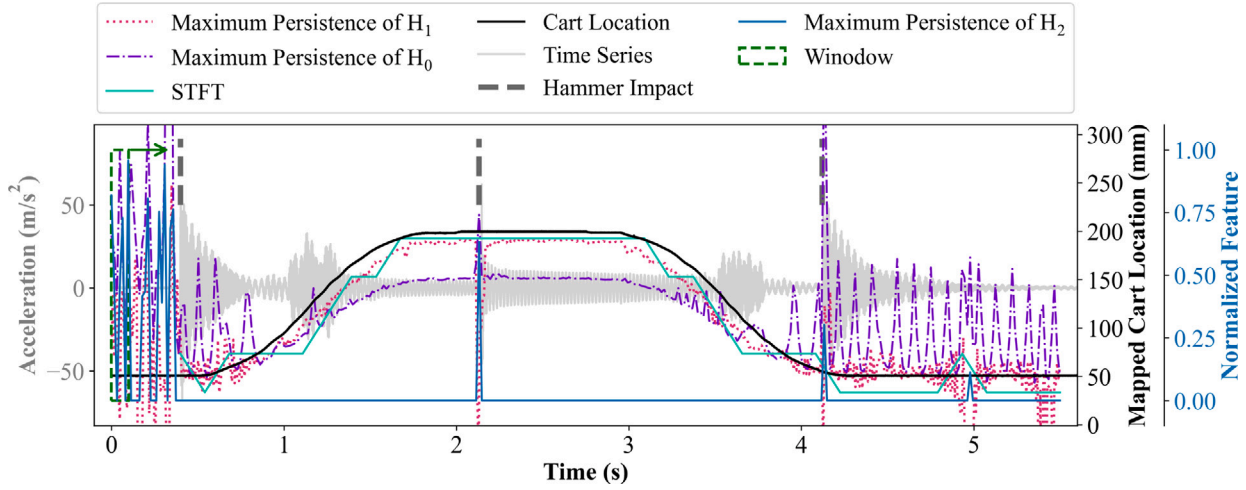

Fig. 6. Evolution of TDA features over test 2 (with impact hammer).

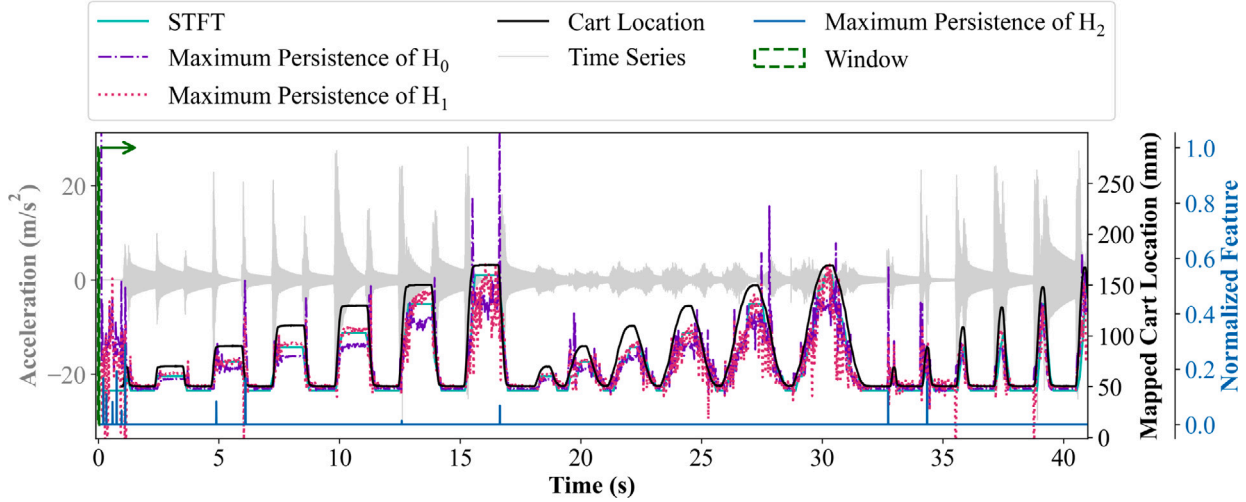

Fig. 7. Evolution of TDA features over test 3.

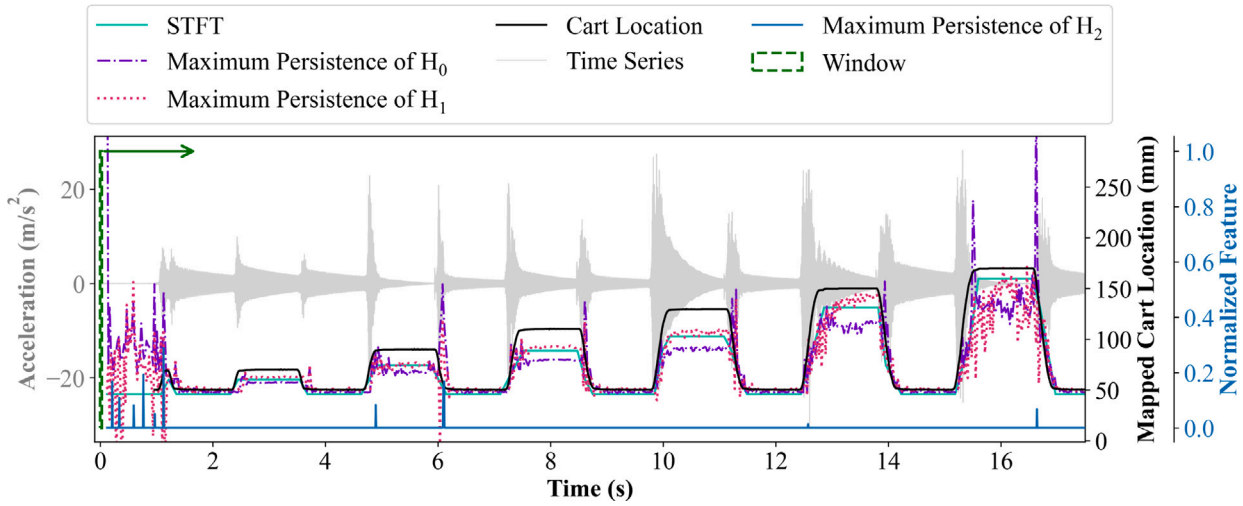

Fig. 8. Evolution of TDA features over test 3 — section 1.

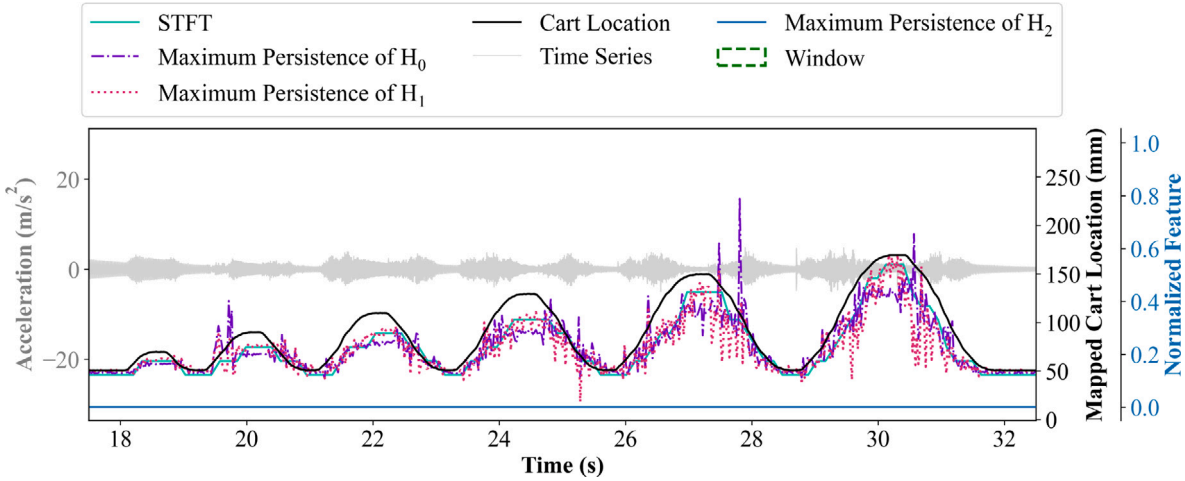

Fig. 9. Evolution of TDA features over test 3 — section 2.

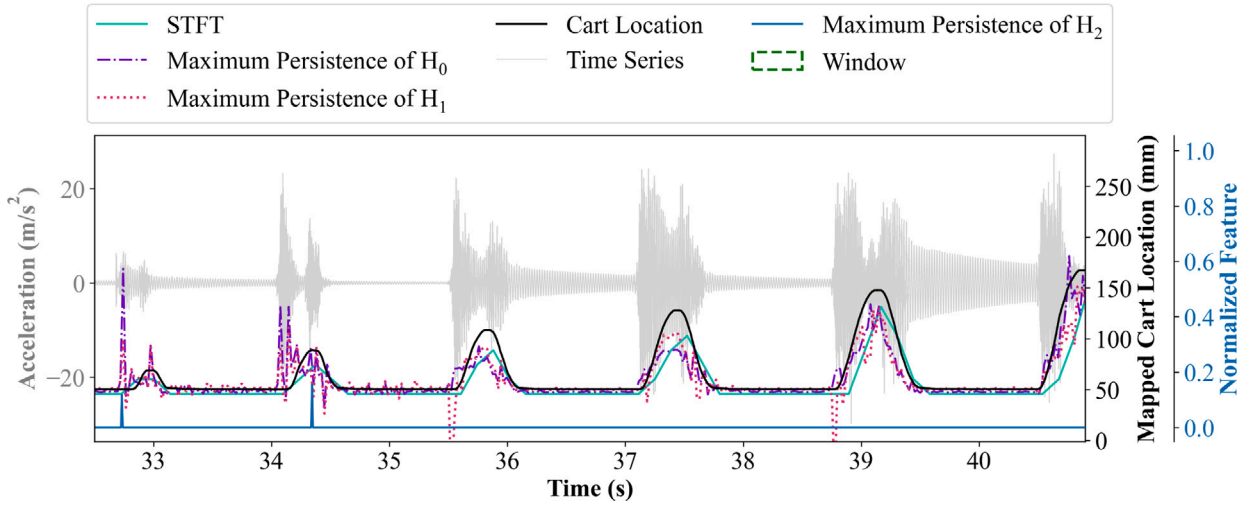

Fig. 10. Evolution of TDA features over test 3 — section 3.

Table 1
Results from LSE.

Test	Feature	a_0	a_1	R^2
Test 1	H_0	−0.41	3.17	0.95
	H_1	−0.8	1.85	0.95
	STFT	−0.13	1.07	0.98
Test 2	H_0	−0.07	5.25	0.63
	H_1	−1.05	2.31	0.91
	STFT	−0.11	1.059	0.96
Test 3	H_0	−0.15	4.25	0.86
	H_1	−0.92	2.47	0.87
	STFT	−0.034	0.93	0.95

Table 2
Performance results for cart localization.

Test	Feature	J_1 (mm)	$J_{2,5}$ (%)	$J_{2,10}$ (mm)	$J_{2,20}$ (%)	$J_{3,5}$ (mm)	$J_{3,10}$ (%)	$J_{3,20}$ (mm)
Test 1	H_1	14.1	65	50	25	2.2	3.8	7.5
	STFT	11.9	81	48	14	2.8	5.6	10.6
Test 2	H_1	16	88	68	18	2.4	5.6	10.5
	STFT	12.1	84	52	14	2.4	5.6	10.8
Test 3 (sect. 1)	H_1	8.7	46	29	12	1.4	2.9	4.9
	STFT	6.3	38	13	3	3.4	4.6	5.4
Test 3 (sect. 2)	H_1	10.2	56	34	15	1.9	3.8	6.0
	STFT	7.3	50	23	6	3.5	4.6	6.4
Test 3 (sect. 3)	H_1	5.7	30	18	8	1.3	2.1	3.4
	STFT	5.9	21	12	5	4	4.4	5.1

absolute error of the cart location when the estimated cart position is within the given distance i (in mm). In other words, J_1 is an overall indication of cart location fitting, $J_{2,i}$ is an indication of how often the estimation is correct within the given error threshold, and $J_{3,i}$ is an indication of how correct the cart location when the estimation is within the given error threshold. Table 2 and Fig. 11 present results for these performance metrics across the three tests through radar plots.

Metric J_1 shows that the TDA metric remains within acceptable ranges (under 20 mm), and that it tends to perform similarly to the STFT for very rapid changes in movements (test 3 — section 3). Metric J_2 shows that the TDA feature maps usually above 80% of the time to within 20 mm of the correct cart location. However, it significantly underperformed for test 2 under 5 and 10 mm, attributable to the use of impact loads. It also outperformed the STFT feature under tests 1 and 2, and showed excellent tracking over test 3 — section 3. Metric J_3 often shows that both the TDA features performed at least similarly to the STFT feature, if not better. In particular, it significantly outperformed the STFT feature under test 3, demonstrating the promise of the TDA feature at tracking rapidly changing states.

5. Conclusion

The objective of this study was to investigate the applicability of TDA features for conducting real-time state estimation on high-rate systems, characterized by rapidly changing dynamics. In this case, the application of interest was the identification of a moving boundary condition through the assessment of the system's frequency, thus reducing the problem of mapping TDA features to the dominating frequency of the system. Some TDA features were explored on a physical perspective, and their applicability to high-rate state estimation problems was assessed. To cope with the system's nonstationarity, a multi-resolution sliding windowing method was proposed, along with a technique to embed time series data into a point cloud. The method was applied to a set of synthetic data, and to realistic experimental data obtained from the dynamic reproduction of projectile in ballistic environments for advanced research (DROPBEAR) testbed, with one set of data including high-rate variations in the boundary condition.

Results from simulations on synthetic data showed that the maximum persistences of H_0 and H_1 provided a stable estimation of the cart location. Results from the Bottleneck and Wasserstein distances, persistence landscape, and silhouette exhibited a higher level of noise. The use of the maximum persistence was further evaluated on DROPBEAR data. Results from low cart acceleration tests showed that both H_0 and H_1 could be used to track the moving cart, with H_1 outperforming H_0. The maximum persistence of H_2 was shown to be useful in detecting noise created by impact loads. Similar results were observed on the higher cart acceleration dataset. A further investigation of the TDA feature showed that it could be used to track the cart location within an acceptable range and that it performed similarly to or overperformed the STFT feature when the estimation was within a given bound. However, the STFT feature outperformed the TDA feature in terms of being within a given estimation bound more often, except for test 3 — section 3 where results were similar.

Overall, results demonstrated that TDA features can be used to track dynamic systems with a dominating, nonstationary fundamental frequency and that it was performing well for more rapid dynamics, in particular relative to the use of an STFT.

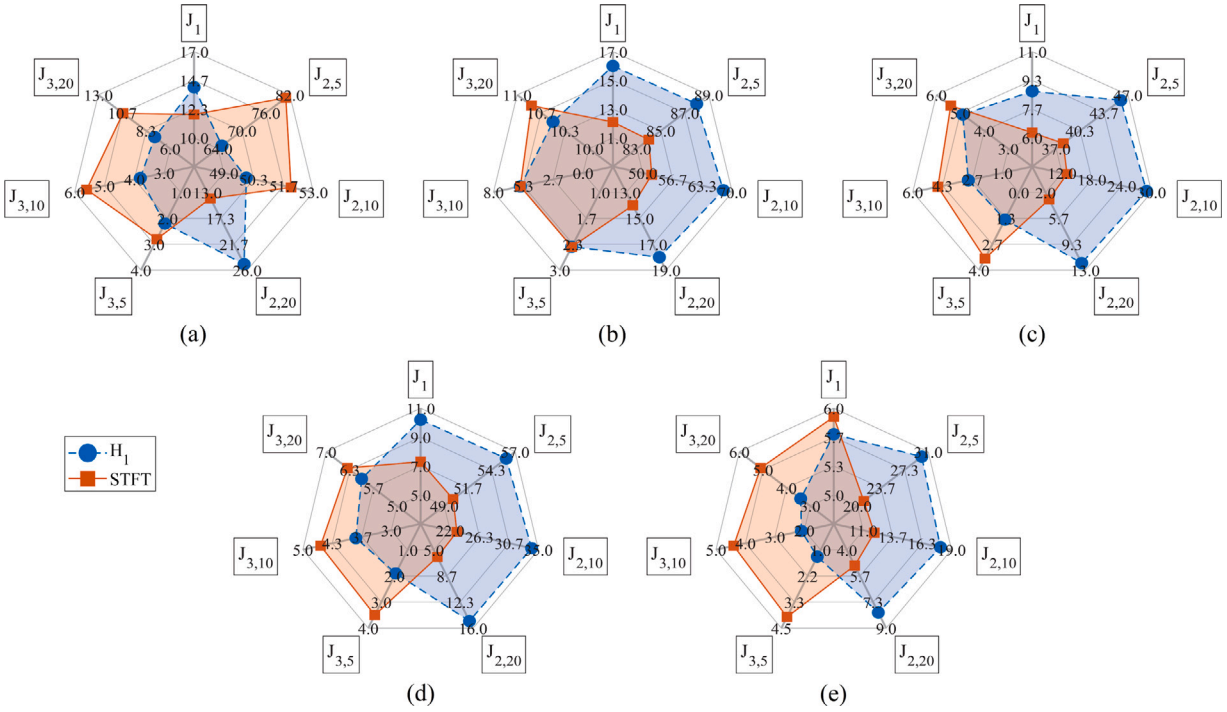

Fig. 11. Radar plots of the performance metrics J_1–$J_{3,20}$: (a) test 1; (b) test 2; (c) test 3-sect. 1; (d) test 3-sect. 2; and (e) test 3-sect. 3.

Future work includes the extension of the study to multi-harmonic systems, and to formulation of algorithms to better map TDA features to the system's state(s).

CRediT authorship contribution statement

Arman Razmarashooli: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yang Kang Chua:** Writing – review & editing, Visualization, Validation, Data curation, Conceptualization. **Vahid Barzegar:** Writing – review & editing, Formal analysis, Conceptualization. **Daniel Salazar:** Writing – review & editing, Visualization. **Simon Laflamme:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Chao Hu:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Austin R.J. Downey:** Writing – review & editing, Supervision, Resources, Data curation. **Jacob Dodson:** Writing – review & editing, Resources, Data curation. **Paul T. Schrader:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the financial support from the Defense Established Program to Stimulate Competitive Research (DEPSCoR), USA award number FA9550-22-1-0303, the Air Force Office of Scientific Research (AFOSR), USA award number FA9550-23-1-0033, and the National Science Foundation, USA awards numbers 1937460 & 2234919. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors, the United States Air Force, or the US government.

Appendix

The following section provides a more formal mathematical definition of TDA. Its content is drawn from known results seen in [45].

Topological Data Analysis (TDA)

Topological Data Analysis offers a framework to study the topology of point clouds using fundamental concepts of modern mathematics, particularly through homology theory [45]. A topological space is a set of points T equipped with a collection of subsets $\mathcal{T}$, which satisfies the following conditions:

- The empty set and the entire set T are included in $\mathcal{T}$.
- The union of any collection of subsets in $\mathcal{T}$ is also in $\mathcal{T}$.
- The intersection of any finite collection of subsets in $\mathcal{T}$ is also in $\mathcal{T}$.

The collection $\mathcal{T}$ is referred to as a topology on T. The subsets in $\mathcal{T}$ are known as open sets. A neighborhood of a point $p \in T$ is any open set that contains p. For example, the open intervals of the real number line $\mathbb{R}$ form a topology on $\mathbb{R}$ often called its standard topology. A metric space X is a specific type of topology that arises from metric d which is a distance function between elements of the set that satisfies the following properties for all $x, y, z \in X$:

- $d(x, y) \geq 0$
- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) + d(y, z) \geq d(x, z)$

An open ball in a metric space (X, d) centered at a point $x \in X$ with radius $r > 0$ is defined as: $B(x, r) = \{y \in X \mid d(x, y) < r\}$. The collection of all such open balls forms a basis for the topology on X. For example, the familiar n-dimensional Euclidean space $\mathbb{R}^n$, where n is a positive integer, is a metric space whose distance function is the standard Euclidean distance. This induces a standard topology based on open balls (recall that the real number line $\mathbb{R} = \mathbb{R}^1$ is also the 1-dimensional version of $\mathbb{R}^n$).

Let $(T, \mathcal{T})$ and $(U, \mathcal{U})$ be topological spaces. A *homeomorphism* is a bijective map $h : T \to U$ such that h and h^{-1} are both continuous. In other words, a homeomorphism is a bijective function h between two topological spaces that preserve the open set structure in both directions [45]. Two topological spaces T and U are said to be homeomorphic if there exists a homeomorphism between them. This relationship defines an equivalence relation on the class of all topological spaces. Hence, two homeomorphic topological spaces are considered topologically equivalent [45].

TDA typically deals with a particular type of topological space called a manifold. An m dimensional manifold M is a topological space where the neighborhoods of all its points are homeomorphic to $\mathbb{R}^m$

There is another notion of comparing topological spaces that is weaker than a homeomorphism since it does not, for example, necessarily preserve dimension. That is known as a homotopy equivalence. Intuitively, it relates how spaces can be continuously deformed from one to the other. Homotopy equivalence is more practical in TDA, showing two spaces are equivalent if one can be continuously deformed from one to the other without cutting or tearing. Let $g : X \to U$ and $h : X \to U$ be continuous maps. A homotopy is a map $H : X \times [0, 1] \to U$ such that $H(., 0) = g(x)$ and $H(., 1) = h(x)$ for all $x \in X$. Two maps g and h are homotopic if there exists a homotopy H connecting them [45]. Two topological spaces T and U are homotopy equivalent if there exist maps $g : T \to U$ and $h : U \to T$ such that $h \circ g$ is homotopic to the identity map $\mathrm{id}_T : T \to T$ and $g \circ h$ is homotopic to the identity map $\mathrm{id}_U : U \to U$. For example, a 2D annulus (or washer) embedded in $\mathbb{R}^2$ is homotopy equivalent to one of its circular 1D boundaries through a linear radial pointwise contraction (a homotopy) of the lines forming the diameter of the annulus's interior to the intersecting points on that annulus's boundary.

To approximate a manifold, simplicial complexes can be used. Simplicial complexes provide a combinatorial framework for studying the topological properties of manifolds. A simplicial complex is a collection of simplices. For $K \geq 0$, a k-simplex σ in an Euclidean space $\mathbb{R}^2$ is the convex hull of a set p of $K + 1$ affinely independent points in R. In particular, a 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, etc.

A geometric simplicial complex K, also known as a triangulation, is a set containing finitely many simplices that satisfies the following two restrictions: (1) K contains every face of each simplex in K. (2) For any two simplices, $\sigma, \tau \in K$, their intersection $\sigma \cap \tau$ is either empty or a face of both σ and τ.

Given a simplicial complex k and a manifold M, we say that K is a triangulation of M if the underlying space $|K|$ is homeomorphic to M. Note that if M is a k-manifold, the dimension of K is also k.

There are many ways to build a simplicial complex. There are two major methods to associate a complex with the point cloud data, namely the Cech Complex and Vietoris–Rips complex [25]. The Čech complex $C_\epsilon(V)$ of a set of points $V = \{v_1, v_2, \ldots, v_n\} \subset \mathbb{R}^d$ at a scale $\epsilon > 0$ is defined as the simplicial complex where a k-simplex $[v_{i_0}, v_{i_1}, \ldots, v_{i_k}]$ is in $C_\epsilon(V)$ if and only if the ϵ-balls centered at these points have a non-empty intersection:

$$[v_{i_0}, v_{i_1}, \ldots, v_{i_k}] \in C_\epsilon(V) \iff \bigcap_{j=0}^{k} B_\epsilon(v_{i_j}) \neq \emptyset. \tag{15}$$

Similarly, the Vietoris–Rips complex $\mathcal{R}_\epsilon(V)$ is defined as the simplicial complex where a k-simplex $[v_{i_0}, v_{i_1}, \ldots, v_{i_k}]$ is in $\mathcal{R}_\epsilon(V)$ if and only if the pairwise distances between the points are at most ϵ:

$$[v_{i_0}, v_{i_1}, \ldots, v_{i_k}] \in \mathcal{R}_\epsilon(V) \iff \|v_{i_j} - v_{i_l}\| \leq \epsilon \text{ for all } 0 \leq j < l \leq k. \tag{16}$$

In simpler terms, in the Čech complex, $k+1$ points form a k-simplex if the intersection of their ϵ-balls is non-empty. In the Vietoris–Rips complex, $k+1$ points form a k-simplex if all pairwise distances between the points are at most ϵ. The choice of ϵ significantly affects the resulting simplicial complex.

Vietoris Rips complexes are a powerful tool for understanding the underlying space's homotopy properties, and to construct these, we need the appropriate value of ϵ. However, in practical application, determining the value of ϵ for real data can be challenging, as it often remains unknown. To address this challenge we adopt a flexible approach. We continuously dilate the real-valued ϵ for all time and construct/record an evolving simplicial complex structure for every ϵ. These complexes, representing different scales or levels of connectivity in the data, form a sequence. This sequence of simplicial complexes is known as a filtered simplicial complex or filtration which captures the developing topological features of the data as ϵ varies. By examining the complexes across the range of ϵ values, we gain insights into the data's topological structures at multiple resolutions. Let $\mathcal{R}(V)$ be an abstract Vietoris–Rips complex. A filtration on $\mathcal{R}(V)$ is defined as a sequence of simplicial complexes

$$\emptyset = \mathcal{R}_0(V) \subset \mathcal{R}_1(V) \subset \mathcal{R}_2(V) \subset \cdots \subset \mathcal{R}(V) \tag{17}$$

A given filtration $\mathcal{R}(V)$ at the topological level induces algebraic representations and mappings between those representations. For example, we can represent any $\mathcal{R}_i(V) \subset \mathcal{R}(V)$ for a positive integer i in the filtration by adjacency matrices. These, in turn, have algebraic group structures in their respective dimensions (or rank). Furthermore, the inclusions of the filtration induce linear transformations between those groups. This creates a (linear) algebraic system that encodes the filtration's evolution over time.

With this representation, we are able to analyze the shape of a given manifold by evaluating these so-called homology groups, denoted as $H_k(M)$, where k represents the dimension of homology. The $H_k(M)$ reveal the presence of multidimensional features such as connected components (H_0), loops or holes (H_1), and voids (H_2), and higher dimensional "holes". The ranks of these homology groups are known as Betti numbers, which count the number of such features.

Data availability

Data will be made available on request.

References

[1] Floris Takens, Detecting strange attractors in turbulence, in: David Rand, Lai-Sang Young (Eds.), Dynamical Systems and Turbulence, Warwick 1980, Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, pp. 366–381.
[2] Gunnar Carlsson, Topology and data, Bull. Amer. Math. Soc. 46 (2) (2009) 255–308.
[3] Paul T. Schrader, Topological multimodal sensor data analytics for target recognition and information exploitation in contested environments, in: Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII, Vol. 12547, SPIE, 2023, pp. 114–143.
[4] Jesse J. Berwald, Marian Gidea, Mikael Vejdemo-Johansson, Automatic recognition and tagging of topologically different regimes in dynamical systems, Discontinuity Nonlinearity Complex. 3 (4) (2014) 413–426.
[5] Firas A. Khasawneh, Elizabeth Munch, Chatter detection in turning using persistent homology, Mech. Syst. Signal Process. 70 (2016) 527–541.
[6] Jose A. Perea, Anastasia Deckard, Steve B. Haase, John Harer, SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data, BMC Bioinformatics 16 (1) (2015).
[7] Saba Emrani, Thanos Gentimis, Hamid Krim, Persistent homology of delay embeddings and its application to wheeze detection, IEEE Signal Process. Lett. 21 (4) (2014) 459–463.
[8] Cássio M.M. Pereira, Rodrigo F. de Mello, Persistent homology for time series and spatial data clustering, Expert Syst. Appl. 42 (15–16) (2015) 6026–6038.
[9] Paul Samuel Ignacio, Christopher Dunstan, Esteban Escobar, Luke Trujillo, David Uminsky, Classification of single-lead electrocardiograms: TDA informed machine learning, in: 2019 18th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2019, pp. 1241–1246.
[10] Hunter Dlugas, Electrocardiogram arrhythmia detection with novel signal processing and persistent homology-derived predictors, Data Sci. 7 (1) (2024) 29–53.
[11] Alperen Karan, Atabey Kaygun, Time series classification via topological data analysis, Expert Syst. Appl. 183 (2021) 115326.
[12] Tristan Gowdridge, Nikolaos Dervilis, Keith Worden, On topological data analysis for structural dynamics: an introduction to persistent homology, ASME Open J. Eng. 1 (2022) 011038.
[13] Yuhei Umeda, Time series classification via topological data analysis, Inf. Media Technol. 12 (2017) 228–239.
[14] Nalini Ravishanker, Renjie Chen, An introduction to persistent homology for time series, Wiley Interdiscip. Rev. Comput. Stat. 13 (3) (2021) e1548.
[15] Morgan Byers, Lee B. Hinkle, Vangelis Metsis, Topological data analysis of time-series as an input embedding for deep learning models, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2022, pp. 402–413.
[16] Jonathan Hong, Simon Laflamme, Jacob Dodson, Bryan Joyce, Introduction to state estimation of high-rate system dynamics, Sensors 18 (1) (2018) 217.
[17] Jacob Dodson, Austin Downey, Simon Laflamme, Michael D. Todd, Adriane G. Moura, Yang Wang, Zhu Mao, Peter Avitabile, Erik Blasch, High-rate structural health monitoring and prognostics: An overview, in: Ramin Madarshahian, Francois Hemez (Eds.), Data Science in Engineering, Volume 9, Springer International Publishing, Cham, 2022, pp. 213–217.
[18] Bryan Joyce, Jacob Dodson, Simon Laflamme, Jonathan Hong, An experimental test bed for developing high-rate structural health monitoring methods, Shock Vib. 2018 (1) (2018) 3827463.
[19] Austin Downey, Jonathan Hong, Jacob Dodson, Michael Carroll, James Scheppegrell, Millisecond model updating for structures experiencing unmodeled high-rate dynamic events, Mech. Syst. Signal Process. 138 (2020) 106551.
[20] Zihan Wu, Michael D. Todd, Uncertainty-quantified damage identification for high-rate dynamic systems, in: Data Science in Engineering, Volume 9: Proceedings of the 39th IMAC, A Conference and Exposition on Structural Dynamics 2021, Springer, 2022, pp. 17–20.
[21] Jin Yan, Simon Laflamme, Jonathan Hong, Jacob Dodson, Online parameter estimation under non-persistent excitations for high-rate dynamic systems, Mech. Syst. Signal Process. 161 (2021) 107960.
[22] Vahid Barzegar, Simon Laflamme, Chao Hu, Jacob Dodson, Ensemble of recurrent neural networks with long short-term memory cells for high-rate structural health monitoring, Mech. Syst. Signal Process. 164 (2022) 108201.
[23] Matthew Nelson, Vahid Barzegar, Simon Laflamme, Chao Hu, Austin R.J. Downey, Jason D. Bakos, Adam Thelen, Jacob Dodson, Multi-step ahead state estimation with hybrid algorithm for high-rate dynamic systems, Mech. Syst. Signal Process. 182 (2023) 109536.

[24] Matthew Nelson, Simon Laflamme, Chao Hu, Adriane G. Moura, Jonathan Hong, Austin Downey, Peter Lander, Yang Wang, Erik Blasch, Jacob Dodson, Generated datasets from dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR) testbed, IOP SciNotes 3 (4) (2022) 044401.
[25] Gunnar Carlsson, Mikael Vejdemo-Johansson, Topological Data Analysis with Applications, Cambridge University Press, 2021.
[26] Zachary Alexander, A Topology-Based Approach for Nonlinear Time Series with Applications in Computer Performance Analysis (Ph.D. thesis), University of Colorado at Boulder, 2012.
[27] Herbert Edelsbrunner, David Letscher, Afra Zomorodian, Topological persistence and simplification, Discrete Comput. Geom. 28 (2002) 511–533.
[28] Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal M. Medina-Mardones, Alberto Dassatti, Kathryn Hess, Giotto-tda: A topological data analysis toolkit for machine learning and data exploration, J. Mach. Learn. Res. 22 (39) (2021) 1–6.
[29] Frédéric Chazal, Bertrand Michel, An introduction to topological data analysis: fundamental and practical aspects for data scientists, Front. Artif. Intell. 4 (2021) 108.
[30] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, Stability of persistence diagrams, in: Proceedings of the Twenty-First Annual Symposium on Computational Geometry, 2005, pp. 263–271.
[31] Marian Gidea, Yuri Katz, Topological data analysis of financial time series: Landscapes of crashes, Phys. A 491 (2018) 820–834.
[32] Herbert Edelsbrunner, John L. Harer, Computational Topology: An Introduction, American Mathematical Society, 2022.
[33] Peter Bubenik, Paweł Dłotko, A persistence landscapes toolbox for topological statistics, J. Symbolic Comput. 78 (2017) 91–114.
[34] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Stochastic convergence of persistence landscapes and silhouettes, in: Proceedings of the Thirtieth Annual Symposium on Computational Geometry, 2014, pp. 474–483.
[35] Jaroslav Stark, Delay embeddings for forced systems. I. Deterministic forcing, J. Nonlinear Sci. 9 (1999) 255–332.
[36] Jaroslav Stark, David S. Broomhead, Michael Evan Davies, J. Huke, Delay embeddings for forced systems. II. Stochastic forcing, J. Nonlinear Sci. 13 (2003) 519–577.
[37] Victoria Caballero, On an embedding theorem, Acta Math. Hungar. 88 (2000) 269–278.
[38] Ziyu Jia, Youfang Lin, Yunxiao Liu, Zehui Jiao, Jing Wang, Refined nonuniform embedding for coupling detection in multivariate time series, Phys. Rev. E 101 (6) (2020) 062113.
[39] Michael Small, Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance, vol. 52, World Scientific, 2005.
[40] Jose A. Perea, John Harer, Sliding windows and persistence: An application of topological methods to signal analysis, Found. Comput. Math. 15 (2015) 799–838.
[41] Michał Adamaszek, Henry Adams, The Vietoris–Rips complexes of a circle, Pacific J. Math. 290 (1) (2017) 1–40.
[42] Michał Adamaszek, Henry Adams, Samadwara Reddy, On Vietoris–Rips complexes of ellipses, J. Topol. Anal. 11 (03) (2019) 661–690.
[43] Eduardo Kausel, Advanced Structural Dynamics, Cambridge University Press, 2017.
[44] Austin Downey, Jonathan Hong, Jacob Dodson, Michael Carroll, James Scheppegrell, Dataset-2-DROPBEAR-acceleration-vs-roller-displacement, 2020.
[45] Tamal Krishna Dey, Yusu Wang, Computational Topology for Data Analysis, Cambridge University Press, 2022.