Topological Data Analysis for Real-Time Extraction of Time Series Features

ARMAN RAZMARASHOOLI, YANG KANG CHUA,
VAHID BARZEGAR, HAN LIU, SIMON LAFLAMME,
CHAO HU, AUSTIN R. J. DOWNEY and JACOB DODSON

ABSTRACT

Real-time state estimation is critical in rapidly assessing structural health to empower real-time feedback mitigation strategies. Of interest to this paper is the state estimation of high-rate system dynamics. High-rate systems are defined as dynamic systems experiencing high-rate (< 100 ms) and high-amplitude (acceleration > 100 g_n) events. Examples include hypersonic vehicles and active impact mitigation strategies. The advanced operation of these mechanisms can only be achieved through control and feedback systems capable of operating in the sub-millisecond range, thus necessitating tight performance constraints. Additionally, high-rate system dynamics are highly nonlinear and non-stationary, for which traditional real-time inference methods cannot provide accurate predictions. Topological data analysis (TDA) is gaining popularity for classifying complex time series. Its integration with architected machine learning algorithms shows promise in advancing the predictive capabilities for high-rate systems. This paper investigates the use of TDA features in conducting state estimation. Some TDA features are explored on a physical perspective, and their applicability to the high-rate state estimation problem is assessed. A promising TDA feature is selected, namely the maximum persistence of H_0, and applied to laboratory datasets extracted from the dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR) testbed. The task consists of detecting the location of a fast-moving boundary condition on a cantilever beam. Results demonstrate that the feature can be used to detect the location of the moving boundary condition online. A discussion on real-time location of a fast-moving boundary condition on a cantilever beam and the applicability to high-rate systems is provided.

Arman Razmarashooli, Ph.D student, Email: shooli@iastate.edu. Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, 50011, USA
Yang Kang Chua, Ph.D student, Email:yang_kang.chua@unconn.edu. Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA
Vahid Barzegar, Ph.D student, Email: barzegar@iastate.edu. Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, 50011, USA

INTRODUCTION

High-rate dynamic systems are defined as systems experiencing dynamic events with amplitudes higher than 100 g_n over a duration of less than 100 ms [1]. Examples of high-rate systems include active blast mitigation systems, advanced weaponry, automotive airbag deployment mechanisms, and hypersonic vehicles [2]. These systems could benefit from real-time feedback capabilities to preserve structural integrity and reliable operations, for example enabling adaptive guidance and active mitigation systems. Yet, achieving real-time state estimation for high-rate dynamic systems remains challenging because of the important time constraint, often under 1 ms, combined with the large levels of complexities and uncertainties characterizing their dynamics [3].

Topological Data Analysis (TDA) is a mathematical and computational framework that applies concepts from algebraic topology to extract information and analyze high dimensional and complex datasets [4]. TDA techniques combine statistical, computation, and topological methods to identify shape-like structures in data that are not apparent through conventional data analysis methods. It is often used to identify and extract the underlying data structure and enable dimensionality reduction [5]. TDA is seen as an improvement with respect to traditional techniques based on algebraic topology, in particular for nonlinear data sets in noisy environments. TDA has gained significant attention in experimental and engineering science [6–9], and its integration with architected machine learning algorithms has shown promise for time series classification [10].

Prior work on real-time state estimation for high-rate systems demonstrated that the application of algebraic topology concepts, in particular, the embedding theorem [11], can be promising in accelerating algorithmic capabilities [12, 13]. This is due to capturing the essential dynamics in the form of delay vectors, which are used to feed machine learning algorithms with minimal information, thus yielding lean and efficient representations. In this paper, we investigate if TDA features can serve a similar function, i.e., used as inputs to an adaptive representation to provide time series forecast capabilities.

Our application problem of interest is the real-time estimation of a moving boundary condition on a cantilever beam. We chose this problem because 1) experimental

Han Liu, Ph.D student, Email: liuhan@iastate.edu. Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, 50011, USA
Simon Laflamme, Professor, Email: laflamme@iastate.edu. Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, 50011, USA
Chao Hu, Associate Professor, Email: chao.hu@uconn.edu. Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA
Austin R.J. Downey, Assistant Professor, Email: austindowney@sc.edu. Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29208, USA
Jacob Dodson, Senior Research Mechanical Engineer, Email: jacob.dodson@eglin.af.mil. Air Force Research Laboratory, Munitions Directorate, Eglin Air Force Base, FL 32542, USA

datasets have been produced on a testbed termed dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR) and have been made widely available to the field, and 2) DROPBEAR datasets are well suited for validating and benchmarking performance of high-rate algorithms [14]. First, we examine some of the available TDA features from a physical perspective to understand how they can relate to the DROPBEAR problem and discuss why we selected the maimixum persistence of the 1-dimensional hole feature (known as H_1) as our key feature of interest. Second, we investigate how point clouds from time series data can be effectively constructed to extract meaningful values for H_1 and demonstrate performance on a synthetic dataset and then on DROPBEAR. Third, conclusions are drawn on the promise of H_1 to serve as an input to a representation built for time series prediction, and a discussion is provided on adapting TDA techniques to the high-rate problem.

TOPOLOGICAL DATA ANALYSIS

The fundamental theory of TDA is based on algebraic topology, where the topological space of a dynamic system is studied using modern mathematical tools, for instance through the evaluation of a point cloud's simplicial complexes [15]. A popular technique to perform TDA is through persistent homology which consists of inspecting the homology groups of a sequence of subspaces for a given dataset. These subspaces are defined by a filtration function such as the Vietoris-Rips complex. As the filtration parameter varies, the homology groups of the subspaces change, and new topological features may appear or disappear [16]. Persistent homology tracks these changes by associating each topological feature with a birth and death time. The persistence of a topological feature is defined as the difference between its death and birth times [17]. The output of persistent homology is a collection of all topological features that can be represented by birth and death times in the form of a persistence diagram [18]. These diagrams can be used to extract several TDA features, including bottleneck distance - a metric that measures the similarity of two persistence diagrams, the number of connected components and loops - a metric that provides insight into the overall structure of the data set like the data sets are fragmented or noisy, and clusterability - an analysis of the distribution of homology classes in the data [19].

To discuss TDA features on a physical context, we first look at our dynamics of interest. DROPBEAR is a cantilever beam subjected to a fast moving boundary condition (i.e., a moving cart or "clamp"). Its dynamics is either excited by the moving cart alone, or also with an impact load using an impact hammer. Thus, the beam is generally undergoing free vibrations, and under that context, we investigate the meaning of TDA features for single-harmonic time series. Consider the equation of motion of a single-degree-of-freedom (SDOF) cantilever beam under free vibration without damping:

$$x(t) = A\cos(\omega t) = \cos(2\pi f t) \quad (1)$$

where ω is the natural circular frequency in units of radian per second, A is the amplitude, $x(t)$ is the motion, f is the natural cyclic frequency in Hz, and t is time in seconds. The first step in conducting TDA is transforming $x(t)$ into a point cloud. This is generally done based on the embedding theorem [11, 20] by embedding the signal into a delay

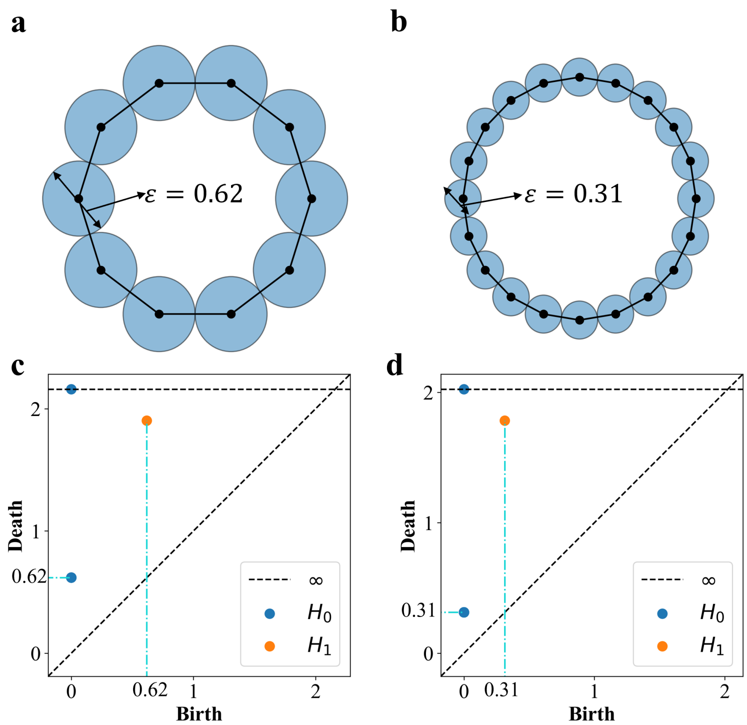

Figure 1. The influence of point cloud density on birth time: (a) Vietoris-Rips complex with 10 points and (b) 20 points; (c) persistence diagram for 10 points; and (d) 20 points. The horizontal dotted line represents infinity.

vector $\chi(t) = [x(t),\ x(t-\tau),\ x(t-2\tau),\ \cdots,\ x(t-(d-1)\tau)]$ where τ is the time delay and d the dimension of the embedding. In the case of a single harmonic signal, prior work has suggested that $d = 2$ is an optimal embedding dimension [21], thus giving raise to a 2-dimensional point cloud with its associated persistence diagram containing information about connected components or zero-dimensional holes (TDA feature H_0) and loops or one-dimensional holes (TDA feature H_1).

As an example, Figures 1(a-b) illustrates two point clouds constructed with a different number of points, and Figures 1(c-d) is a plot of their associated persistence diagrams. A circle of diameter ε is drawn around each point until the points are connected. When the points are all connected and form a "hole", H_1 is born. After, ε is increased until the hole vanishes, with corresponds to H_1's death time. In this noise-free example, H_1 is born when the first H_0 dies (the second H_0 represents the feature with infinite persistence). Remark that the birth time of H_1 decreases and approaches 0 with an increased sampling rate, and its death time converges to $\sqrt{3}$. This can be observed by visually comparing H_1 between Figures 1(c-d).

In contrast to the first homology group (H_1), which relates to the frequency of the harmonic signal, the topological feature known as the zero-dimensional hole (H_0) lacks physical significance when analyzed in the context of harmonic signals. This is because for a dense circular point cloud, the death time of the zero-dimensional hole will converge to zero as the number of points increases, making it a non-informative feature. Similarly, the number of connected components in the point cloud merely indicates the sampling rate, and, in the absence of noise, the number of H_1 features is always one. Accordingly, without noise, the bottleneck distance metric relies exclusively on the maximum persistence of H_1 and does not yield any significant insights from a physical perspective. As a result, we select the maximum persistence H_1 as the feature of interest and study its applicability to our system of interest.

In further studying of H_1, one can notice that the reconstructed state space of harmonic signals for the time delay equal to $0.25/f$ is always a unit circle and that the maximum persistence of this circle is equal to $\sqrt{3}$ and is independent of the frequency of the signal [22]. However, for a time delay below the optimum $(0.25/f)$, the reconstructed state space will be an ellipse, and the ratio of its major axis to its minor axis determines the maximum persistence of H_1. More circular ellipse results in higher maximum persistence of H_1 [23,24]. Hence, under a fixed time delay τ, the frequency of the harmonic signal will affect the ratio of the major axis to the minor axis, and this, in turn, will affect the maximum persistence of H_1. It follows that, from a physical perspective, the maximum persistence of H_1 can be a useful feature in classifying harmonic signals in terms of their frequencies.

A key remark is that this interpretation of H_1 as a feature for harmonic signals only applies for stationary systems. Our dynamics of interest is non-stationary, because of the moving boundary condition. To cope with the problem of non-stationarity, the strategy will be to apply a sliding window over the dataset to extract local values for H_1.

Windowing for Non-Stationary Signal

The embedding theorem states that $\chi(t)$ is topologically equivalent to $x(t)$ and thus can be used to reconstruct $x(t)$ given a function g with $x(t) = g(\chi(t))$, for $\chi(t)$ constructed with appropriate values of τ and d. Here, we are not necessarily focused on reconstructing $x(t)$, but on finding TDA features that relate to its dynamics. Yet, it is important to know that, for a harmonic signal, these values for τ and d are $0.25/f$ and $d = 2$. In coping with a non-stationary signal, we set a maximum allowable $\tau = 0.25/f_{\text{max}}$ with f_{max} the maximum frequency of the system. The use of a higher τ would risk of folding the topological space onto itself, and information could be lost. As briefly introduced above, the time series is inspected using a sliding window of size $1/f_{\text{min}} + \tau$ with f_{min} the minimum frequency of the system to ensure the point cloud will form a complete loop. This approach ensures that the window size is sufficiently large to capture the characteristics of all frequencies in the data. One must also ensure that τ not be unreasonably small and be selected to allow the phase space to sufficiently unfold and thus generate topological features.

METHODOLOGY

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

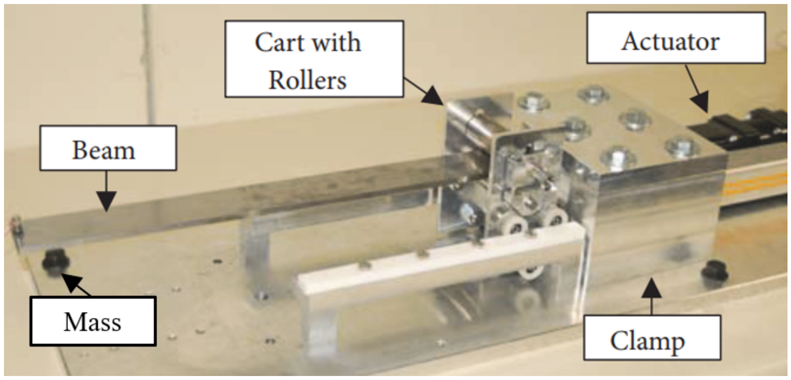

Figure 2. Picture of the DROPBEAR testbed. [25]

Case Study 1: Synthetic Harmonic Signal

The signal is taken as

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

with $f(t)$ varying between 1 and 3 Hz. The excitation is plotted in Figure 3(a) (red solid line). In this excitation, the frequency remains constant at 1 Hz for the first two seconds (0-2 seconds) before increasing to 3 Hz over the next two seconds (2-4 seconds), after which it remains at 3 Hz for another 2 seconds. The size of the moving window is $1/f_{\text{min}} + \tau = 1 + 0.03 = 1.03$ seconds, with data embedded using $\tau = 0.03$ seconds ($0.25/f_{\text{max}} = 1/12$ seconds).

Case Study 2: DROPBEAR

The DROPBEAR testbed was designed and developed to validate state estimation algorithms for high-rate dynamic systems. Briefly, the testbed (Figure 2) consists of a 505 mm cantilever steel beam with a mass attached at its tip using an electromagnet, a PCB 353B17 accelerometer installed 300 mm away from the clamp, and a sliding cart that can be moved along the beam with a linear actuator. The electromagnet is used to simulate a real-time loss in the system's mass. However, in our study, the mass remains attached during the movement of the cart. The dataset is taken from the dataset made available online, corresponding to Dataset-6, Test 9 [26]. The cart is initially located 50 mm from the clamp, moves at 200 mm from the clamp, and comes back its original position. The temporal location of the cart is plotted in Figure 3(b) (back 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 embedded using $\tau = 0.004$ seconds ($0.25/f_{\text{max}} = 0.008$ seconds) and the size of the moving window is was taken as $1/f_{\text{min}} + \tau = 0.06$ seconds which was used over the time series data. The window is slid every 0.001 seconds to reduce computation demands.

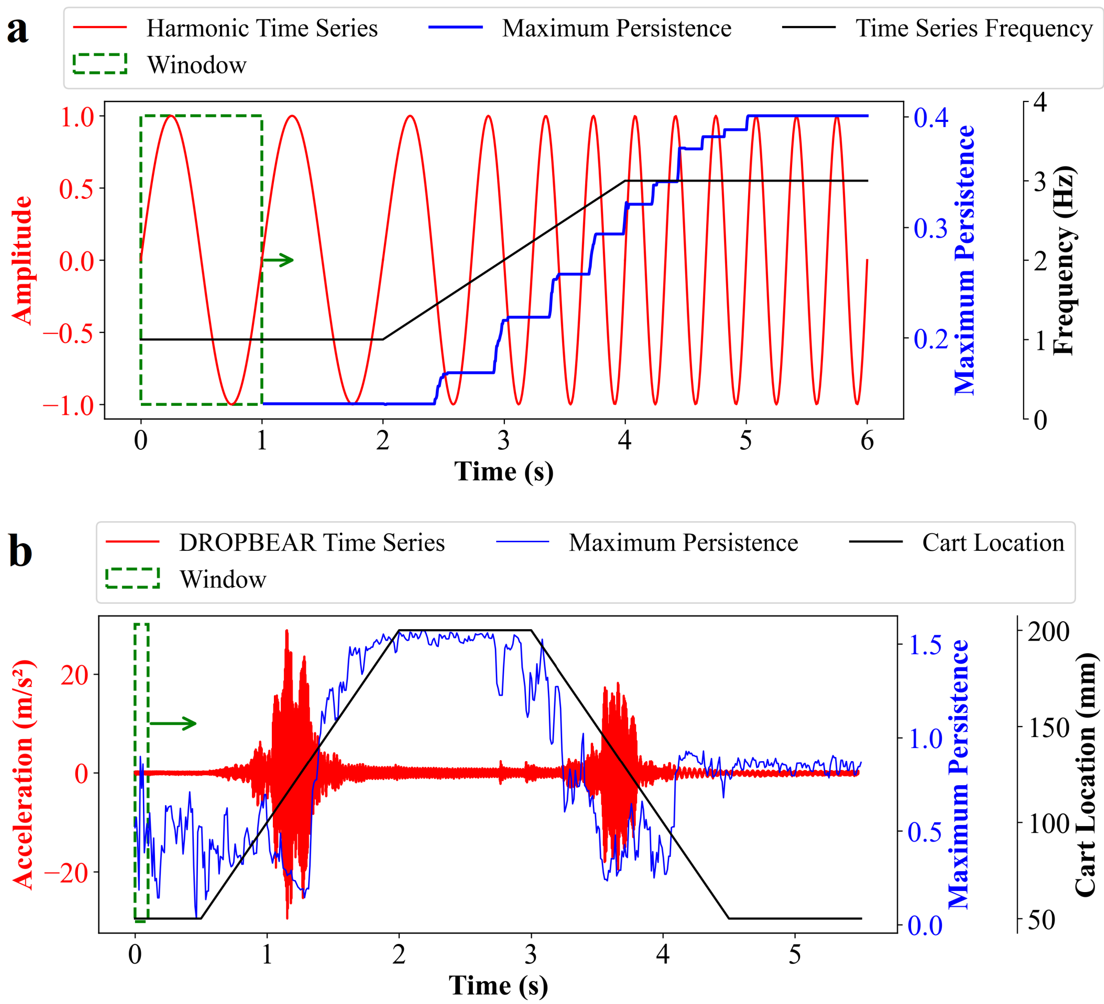

Figure 3. (a) Changes of the maximum persistence in a generated synthetic sine signal; and (b) change of the maximum persistence in the DROPBEAR test.

RESULTS AND DISCUSSION

Results are presented in this section, followed by a discussion on high-rate applicability.

Case Study 1: Synthetic Harmonic Signal

Figure 3(a) shows how the maximum persistence of H_1 evolves against the harmonic frequency (see the solid blue line). It can be observed that the maximum persistence remains constant at the beginning when the frequency of the signal is constant (at 1 Hz), and starts increasing when the frequency changes to eventually plateau when the frequency stabilizes again (at 3 Hz). This results demonstrates that H_1 directly relates to the system's frequency, but that an exact mapping during a change in dynamics could be more difficult to achieve because of the window size overlapping many different frequencies.

Case Study 2: DROPBEAR

Figure 3(b) plots the time series acceleration signal from the sliding cart without hammer excitations, and compares it against the maximum persistence of H_1. Results show that here too, the TDA feature relates to the system's frequency and thus the cart location. However, H_1 appears sensitive to this noisy environment, and also suffers from the lag provoked by the window size, let along computation time that is out-of-the-scope of this paper.

Discussion

The objective of this study was to investigate the applicability of TDA features for conducting real-time state estimation. 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 frequency of a single harmonic. While we demonstrated that the maximum persistence of H_1 can be used for the task, other features could be of interest, including those that can be extracted through other TDA tools, such as persistent cohomology. A key limitation of the high-rate problem is the computation speed. In fact, constructing persistence diagrams is highly time-consuming. In our example on the DROPBEAR data, the construction of each persistence diagram took 2.877 seconds, far above the sub-millisecond requirement. It follows that different methods should be developed to conduct fast TDA feature extraction, both from a hardware and software perspective. The present technique also relies on some level of physical knowledge from the system of interest, for instance, a bound on the dominating frequencies to construct the window size and time delay. In addition, true high-rate systems, while also typically undergoing free vibrations after impact, have dynamics that are far more complex, and thus the embedding dimension d is expected to be larger, giving rise to more TDA features (i.e., holes up to H_{d-1}). Overall, the demonstration produced in this research shows that some TDA features embed important information about the system's dynamics and could be leveraged to construct state estimation and time series forecasting representations.

CONCLUDING REMARKS

This paper presented a preliminary study on the use of TDA features in conducted state estimation of high-rate system dynamics. The problem of interest was the detection of non-stationary harmonic frequency. Some TDA features were explored on a physical perspective, and their applicability to the high-rate state estimation problem was assessed. A promising TDA feature was selected, namely the maximum persistence of H_1. To cope with the system's non-stationarity, a windowing method was proposed, along with a technique to embed time series data into a point cloud. The method was applied to both synthetic data and laboratory data obtained from the dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR) testbed. Results show that the TDA feature could map to the system's state. However, more investigations and refinements are required to improve performance in a noisy environment (i.e., DROPBEAR).

While showing important promise, TDA features are difficult to apply to the high-rate problem. The computation time required in producing persistence diagrams is the main bottleneck. Different methods will need to be developed to conduct fast TDA feature extraction, both on a hardware and software perspective. This is part of our future work.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support from the Defense Established Program to Stimulate Competitive Research (DEPSCoR) award number FA9550-22-1-0303 and the Air Force Office of Scientific Research (AFOSR) award number FA9550-23-1-0033. 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 United States Air Force.

REFERENCES

1. Dodson, J., B. Joyce, J. Hong, S. Laflamme, and J. Wolfson. 2017. "Microsecond State Monitoring of Nonlinear Time-Varying Dynamic Systems," in *Volume 2: Modeling, Simulation and Control of Adaptive Systems Integrated System Design and Implementati on Structural Health Monitoring*, American Society of Mechanical Engineers, doi:10.1115/smasis2017-3999.
2. Dodson, J., A. Downey, S. Laflamme, M. D. Todd, A. G. Moura, Y. Wang, Z. Mao, P. Avitabile, and E. Blasch. 2021. "High-Rate Structural Health Monitoring and Prognostics: An Overview," in *Data Science in Engineering, Volume 9*, Springer International Publishing, pp. 213–217, doi:10.1007/978-3-030-76004-5_23.
3. Hong, J., S. Laflamme, J. Dodson, and B. Joyce. 2018. "Introduction to State Estimation of High-Rate System Dynamics," *Sensors*, 18(2):217, doi:10.3390/s18010217.
4. Carlsson, G. E. 2009. "Topology and data," *Bulletin of the American Mathematical Society*, 46:255–308.
5. Wasserman, L. 2018. "Topological Data Analysis," *Annual Review of Statistics and Its Application*, 5(1):501–532, doi:10.1146/annurev-statistics-031017-100045.
6. Carrière, M., S. Y. Oudot, and M. Ovsjanikov. 2015. "Stable Topological Signatures for Points on 3D Shapes," *Computer Graphics Forum*, 34(5):1–12, doi:10.1111/cgf.12692.
7. Skraba, P., M. Ovsjanikov, F. Chazal, and L. Guibas. 2010. "Persistence-based segmentation of deformable shapes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 45–52, doi:10.1109/CVPRW.2010.5543285.
8. Offroy, M. and L. Duponchel. 2016. "Topological data analysis: A promising big data exploration tool in biology, analytical chemistry and physical chemistry," *Analytica Chimica Acta*, 910:1–11, doi:10.1016/j.aca.2015.12.037.
9. Skaf, Y. and R. Laubenbacher. 2022. "Topological data analysis in biomedicine: A review," *Journal of Biomedical Informatics*, 130:104082, doi:10.1016/j.jbi.2022.104082.
10. Ravishanker, N. and R. Chen. 2019, "Topological Data Analysis (TDA) for Time Series," doi:10.48550/ARXIV.1909.10604.
11. Takens, F. 1981. "Detecting strange attractors in turbulence," in *Lecture Notes in Mathematics*, Springer Berlin Heidelberg, pp. 366–381, doi:10.1007/bfb0091924.

12. Barzegar, V., S. Laflamme, C. Hu, and J. Dodson. 2022. "Ensemble of recurrent neural networks with long short-term memory cells for high-rate structural health monitoring," *Mechanical Systems and Signal Processing*, 164:108201, doi:10.1016/j.ymssp.2021.108201.
13. Yan, J., S. Laflamme, J. Hong, and J. Dodson. 2021. "Online parameter estimation under non-persistent excitations for high-rate dynamic systems," *Mechanical Systems and Signal Processing*, 161:107960, doi:10.1016/j.ymssp.2021.107960.
14. Nelson, M., S. Laflamme, C. Hu, A. G. Moura, J. Hong, A. Downey, P. Lander, Y. Wang, E. Blasch, and J. Dodson. 2022. "Generated datasets from dynamic reproduction of projectiles in ballistic environments for advanced research (DROPBEAR) testbed," *IOP SciNotes*, 3(4):044401, doi:10.1088/2633-1357/aca0d2.
15. Edelsbrunner, H. and J. Harer. 2010. *Computational Topology - an Introduction.*, American Mathematical Society, ISBN 978-0-8218-4925-5.
16. Aktas, M. E., E. Akbas, and A. E. Fatmaoui. 2019. "Persistence homology of networks: methods and applications," *Applied Network Science*, 4(1), doi:10.1007/s41109-019-0179-3.
17. Carlsson, G. and M. Vejdemo-Johansson. 2021. *References*, Cambridge University Press, p. 207–215, doi:10.1017/9781108975704.011.
18. Edelsbrunner, H., D. Letscher, and A. Zomorodian. 2000. "Topological persistence and simplification," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pp. 454–463, doi:10.1109/SFCS.2000.892133.
19. Hatcher, A. 2002. *Algebraic topology*, Cambridge University Press, Cambridge, ISBN 0-521-79160-X; 0-521-79540-0.
20. Kantz, H. and T. Schreiber. 2003. *Nonlinear Time Series Analysis*, Cambridge University Press, ISBN 9781139440431.
21. Small, M. 2005. *Applied nonlinear time series analysis: applications in physics, physiology and finance*, vol. 52, World Scientific.
22. Perea, J. and J. Harer. 2013, "Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis," doi:10.48550/ARXIV.1307.6188.
23. Adamaszek, M. and H. Adams. 2015. "The Vietoris-Rips complexes of a circle," *Pacific Journal of Mathematics*, 290, doi:10.2140/pjm.2017.290.1.
24. Adamaszek, M., H. Adams, and S. Reddy. 2017. "On Vietoris-Rips complexes of ellipses," *Journal of Topology and Analysis*, 11, doi:10.1142/S1793525319500274.
25. Joyce, B., J. Dodson, S. Laflamme, and J. Hong. 2018. "An Experimental Test Bed for Developing High-Rate Structural Health Monitoring Methods," *Shock and Vibration*, 2018:1–10, doi:10.1155/2018/3827463.
26. High-Rate SHM Working Group. 2022, "Dataset 6 - DROPBEAR Data," `https://github.com/High-Rate-SHM-Working-Group/Dataset-6-DROPBEAR_data`, accessed: 3.1.2023.