Using Topological Data Analysis to Visualize Instrument Output

S.N. Chukanov[1,A], I.S. Chukanov[2,B]

A Sobolev Institute of Mathematics of the Siberian Branch of RAS,
Omsk branch, Omsk, Russia
B Ural Federal University named after the first President of Russia B. N. Yeltsin,
Ekaterinburg, Russia

[1] ORCID: 0000-0002-8106-9813, ch_sn@mail.ru
[2] ORCID: 0000-0001-9946-7484, chukanov02@gmail.com

Abstract

The article discusses Takens embedding for 2D and 3D visualization of 1D time series data. The paper considers the use of topological data analysis in conjunction with Takens embedding to analyze the output information of a dynamic system - a rigid body. 3D images of curves constructed from three components of the angular velocity vector of a rigid body for various values of the main moments of inertia are constructed. Three-dimensional images of the curves constructed on the basis of the Takens embedding for the component of the angular velocity of a rigid body for various values of the principal moments of inertia are constructed. Distances between 3D images of curves are determined by constructing persistent landscape functions. Using the method of topological data analysis in conjunction with Takens embedding for image comparison allows you to classify and identify images and output information of the instrumental composition.

Keywords: topological data analysis, persistent homology, Takens embedding, rigid body dynamics.

1. Introduction

The article deals with Takens embedding for 2D and 3D visualization of 1D time series data. Topological data analysis (TDA) is a field of science in which the topological properties of data are analyzed. In recent years, interest has increased in the use of TDA methods [1, 2, 3] and application in various fields of knowledge. TDA assumes that the data is in a form that can make a difference. Early contributions to the field of TDA were made by Edelsbrunner H. and Harer J. L. [1]. Zomorodian A. and Carlsson G. used the basis for the development of the TDA methodology: Persistent Homology [2, 3, 4]. The goal of TDA is to define informative topological properties and use them as descriptors.

The key mathematical tool in topological data analysis is the persistent homology (PH) method [1, 4], which is used to extract topological information from data. Consider a way to form PH from data points in Euclidean space. The goal is to derive the topology from the final data. Consider r-balls (of radius r) for topology reconstruction. It is expected that the r-ball model can represent the main topological structures. If r is small, then the union of all r-balls consists of non-intersecting r-balls. If the radii r are too large, then the union becomes one spatial component. Persistent homology considers all values r at the same time and provides an expression for topological properties.

The TDA method can be used to extract knowledge from time series [5, 6, 7]. The states of dynamical systems change in time; in this case, time series are formed. The state of the system x_t at a point in time t is a description of the system, and the evolution of the system in the state space is determined by the transition function Φ. Attractors determine the set of system states, to the points of which the trajectory is directed. n-dimensional manifold is a topological space M, for which each point $x \in M$ has a neighborhood homeomorphic to the

Euclidean space $\mathbf{R}^n$. A smooth map $\Phi : M_1 \to M_2$, where M_1 and M_2 are smooth manifolds, is an embedding of M_1 in M_2 if Φ is diffeomorphism from M_1 to a smooth submanifold of M_2; then M_2 is an embedding space with embedding dimension $\dim(M_2)$. Takens embedding of coordinates [8] allows to transform a time series in a space of higher dimension, so that the topology of the original manifold that generates the values of the time series is preserved. The Takens method finds a function Φ that maps the manifold M_1 into the manifold $M_2 : M_1 \to M_2$, where $\dim(M_2)$ is the dimension of the embedding.

Takens embedding allows you to transform time series data into meaningful point clouds to calculate persistent homology [8, 9, 10, 11]. In addition, the use of the sliding window method allows segmenting long time series into fragments, which makes topological features comparable within and between data sets.

The Takens-embedded TDA method is widely used in chemistry and biological systems [12], signal theory [13], in the study of dynamic systems [14], etc. The use of the TDA method for comparing images [15, 16] makes it possible to classify and identify images (or signals of a different physical nature).

In this paper, we consider the use of TDA together with the Takens embedding to analyze the output information of a dynamical system - a rigid body.

Examples of rigid bodies are objects such as aircraft that contain control systems (eg, attitude control systems). The instrumental composition of control systems consists of sensitive elements and executive bodies. To measure the three components of the angular velocity vector of a rigid body, sensitive elements are usually used - angular velocity sensors (or a block of angular velocity sensors that generates information about the projections of the angular velocity vector of a rigid body on the axes of the main moments of inertia of a rigid body).

3D images of curves constructed from three components of the angular velocity vector of a rigid body for various values of the main moments of inertia are constructed. 3D images of the curves built on the basis of the Takens embedding for the components of the angular velocity ω_y of a rigid body for various values of the main moments of inertia are constructed.

Using the TDA method together with the Takens embedding to compare images (or signals of other physical nature) allows you to classify and identify information from sensors.

2. Topological data analysis for time series analysis

Geometry represented by data in metric space is not always up to date; sometimes more basic properties are of interest, such as the number of components, holes, or voids. Algebraic topology [17] fixes these properties by associating the vector spaces with them. The field coefficient homology assigns to a vector space $H_i(X)$ the space X for each $i \in \{0,1,2,...\}$ such that $\dim(H_0(X))$ is the number of connected components in X, $\dim(H_1(X))$ is the number of holes in X, $\dim(H_2(X))$ is the number of voids in X and $\dim(H_k(X))$ – the k-th homology group in X describes k-dimensional holes in X.

A simplex is the n-dimensional analogue of a triangle or tetrahedron. A n-simplex is an n-dimensional polyhedron created by the convex hull of its $n+1$ vertex. Let σ be n-simplex. A vertex σ is each of the $n+1$ points used to define σ, and a face σ is the convex hull of any subset of vertices σ. A simplicial complex is a topological space realized as the union of any set of simplices Σ with the following two properties: (1) any face of Σ the simplex also lies in Σ; (2) the intersection of any two simplices Σ is also a simplex.

The persistent homology method studies the qualitative aspects of the data by calculating their topological characteristics. It is robust to perturbations, is independent of the size and coordinates of the embedding, and can provide a representation of the qualitative

characteristics of the data. The input data is a point cloud in metric space; for example $X = \{x_1, ..., x_n\}$, in Euclidean space $\mathbf{R}^d$. To compare the topological space, simplicial complexes are constructed.

After the calculation of simplicial complexes, features dominate in the space consisting of vertices, edges and polyhedra of higher dimensions. Then, using homology, characteristics such as components, holes, voids, and other higher-dimensional equivalent characteristics can be measured. The permanence of these functions is represented on persistent diagrams or persistent barcodes.

Persistent homology captures how long topological features persist. The ranks of persistent homologous groups are presented in the persistence diagrams. This is multiset of points in $\mathbf{R}^2$ and is defined as [1]; persistent homology can be visualized by a persistence diagram (PD) $D = \{(b_i, d_i) \in \mathbf{R}^2 \mid i \in I, b_i \leq d_i\}$. Each point $(b_i, d_i) \in D$, which is called a persistent homology generator, represents a topological property that represents a topological property that appears at X_{b_i} and disappears at X_{d_i} in the r-ball model. A topological property with high persistence $\text{pers}_i = d_i - b_i$ can be considered as a reliable structure, while a topological property with low persistence can be considered as noise. Persistent diagrams encode topological and geometric information about data points.

Maximum persistence is defined as:
$$\text{maxPers}(D_i) = \max_{(birth, death) \in D_i} (death - birth),$$
where D_i is the persistence diagram for the i-th homology.

Dynamic systems are built from the state space. The system is considered dynamic because states can change over time. Dynamic systems can be either deterministic or stochastic. The dynamic system is described in the state space X, space of time T and is determined by the transition function $\Phi : X \times T \to X$. The state x_{t+1} is defined through the state x_t using the transition function $\phi(\cdot) \in \Phi$:
$$x_{t+1} = \phi(x_t); \phi(\cdot) \in \Phi,$$
where $t = 0, 1, 2 ... \in T$.

3. Takens embedding for Time Series Data

Dynamic systems are used to model systems whose states change over time. The state x_t at a point in time t is a description of the system, and the evolution of the system in the state space is determined by the transition function Φ. Attractors determine the set of system states, to the points of which the trajectory is directed. The time series is determined by the observed states of the dynamical system.

n-dimensional manifold is a topological space M, for which each point $x \in M$ has a neighborhood homeomorphic to the Euclidean space $\mathbf{R}^n$. A smooth mapping $\Phi : M_1 \to M_2$, where M_1 and M_2 are smooth manifolds, is an embedding M_1 to M_2, if Φ is a diffeomorphism from M_1 to M_2 a smooth submanifold (embedding space).

Takens embedding of coordinates allows you to reconstruct a time series in a higher dimensional space, so that the topology of the original manifold that generates the values of the time series is preserved. Takens suggested that d-dimensional manifold containing an attractor A can be embedded in $\mathbf{R}^{2n+1}$ [8]. The Takens method finds a function Φ that maps $M_1 \to M_2$, where is the nesting dimension $\dim(M_2)$, which can be $\mathbf{R}^{2n+1}$.

Thus, the Takens embedding makes it possible to obtain a continuous transformation of the original manifold M to $X \in \mathbf{R}^d$, where d is the embedding dimension and X is the trajectory matrix. Let $x = \{x_1, x_2, \ldots, x_N\}$ be a time series and X be a trajectory matrix:

$$X = \begin{bmatrix} X_{1+(d-1)\tau} \\ X_{2+(d-1)\tau} \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1+(d-1)\tau} & \cdots & x_{1+\tau} & x_1 \\ x_{2+(d-1)\tau} & \cdots & x_{2+\tau} & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & \cdots & x_{N-(d-1)\tau} & x_{N-(d-1)\tau} \end{bmatrix} \tag{1}$$

where each point in space is represented by a string. A more formal definition is given in [10].

Suppose that $x(t) = v_j(y)$ for some $j = 1, \ldots, n$, where $v(t) = (v_1(t), \ldots, v_n(t))$ is a curve on the manifold Ω. Suppose that $v(t)$ visits every part of Ω, which means that $v(t)$ is dense in Ω with respect to its topology. Then there exists $\tau > 0, K \in Z$, where Z denotes real numbers such that the corresponding vectors $(x(t), x(t+\tau), \ldots, x(t+K\tau))$ are on a manifold topologically equivalent to Ω.

Example 1. Takens embedding example.

Denote by $X_{(k,\tau,m)}$ the reconstructed one-dimensional time series, where k is the time, τ is the delay time, and m is the dimension of the reconstruction. Let our one-dimensional time series be $X = [x_0, \ldots, x_{12}]^T$. Using (1) for a one-dimensional reconstruction with $k = 0, \tau_1 = 1, m_1 = 2$, we obtain a time series $X_{(0,1,2)} = \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \\ \vdots & \vdots \\ x_{11} & x_{12} \end{bmatrix}$.

The definition of the Takens embedding dimension is based on the false nearest neighbors method [9]. The embedding property is that when the embedding dimension m is too small, distant points in the original phase space are close points in the reconstructed phase space. These points are called false neighbors. When calculating the false nearest neighbor for each point x_i, we look for the nearest neighbor x_j in m-dimensional space. After that, the ratio $R_i = \dfrac{|x_{i+1} - x_{j+1}|}{|x_i - x_j|}$ is calculated. If the ratio R_i exceeds the specified threshold R, then the point is marked as a false neighbor. If the attachment size is high enough, the ratio R_i is zero. One way to calculate R_i is to embed the lag time series x with delay τ in a range of different embedding dimensions m. You need to find all nearest neighbors and calculate the percentage of neighbors left after expanding the extra dimensions.

When evaluating the time delay τ, two criteria are important: 1) the time delay τ must be large enough so that the information about the value of x at time $n+\tau$ differs from the information already known from observing the value of x at time n; 2) the time delay τ should not be large enough so that the system does not forget about its initial state [11].

4. Determination of distances between images

To compare images, let's determine the distances between these images: the greater the difference between the images, the greater the distance between them; the distance between

identical images is zero. Euclidean transformations of images should not change the distance between them.

To determine the distance between images (or objects of other physical nature), a persistent landscape is used - a piecewise linear function, which is a generalization of a persistent diagram [18]. The persistent landscape rotates the persistence diagram so that the diagonal becomes the new axis x. The i-th order of persistence landscape produces a piecewise linear function of the i-th largest point value in the persistence diagram after rotation. For the pair $p = (b,d) \in D$, where D are the persistence diagrams, the piecewise linear functions $\Lambda_p(t) : \mathbf{R} \to [0, \infty]$ are equal to:

$$\Lambda_p(t) = \begin{cases} t - b, & t \in \left[b, \dfrac{b+d}{2} \right], \\ d - t, & t \in \left[\dfrac{b+d}{2}, d \right], \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Then the persistent landscape (PL) function is defined as:

$$F : \mathbf{R} \to \mathbf{R} : F(t) = \sup_{p \in D} \left(\Lambda_p(t) \right). \tag{3}$$

Let's form the core of PL functions:

$$K\left(D^{(1)}, D^{(2)} \right) = \left\langle \lambda^{(1)}, \lambda^{(2)} \right\rangle = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \lambda_k^{(1)}(t) \lambda_k^{(2)}(t) \, dt. \tag{4}$$

For the PL functions, we form the p-norm [18]:

$$\|\lambda\|_p = \sum_{k=1}^{\infty} \left[\int_{-\infty}^{\infty} \left(\lambda_k(t) \right)^p \, dt \right]^{\frac{1}{p}}, \tag{5}$$

where $1 \leq p < \infty$.

Distances between PL functions can be determined using the p-norm:

$$\left\| \lambda^{PL} - \lambda'^{PL} \right\|_p = \left[\sum_k \int_{-\infty}^{\infty} \left| \lambda_k^{PL}(t) - \lambda_k'^{PL}(t) \right|^p \, dt \right]^{\frac{1}{p}}, \tag{6}$$

where $1 \leq p < \infty$.

Let X and Y be two persistence diagrams. A persistent diagram consists of a finite number of points above the diagonal. To this finite set we add an infinite number of points on the diagonal. Consider bijections $\eta : X \to Y$ and write sup (least upper bound) of the distances between the corresponding points for each. By measuring the distance between the points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ with the norm:

$$\|x - y\|_\infty = \max \left\{ |x_1 - y_1|, |x_2 - y_2| \right\}, \tag{7}$$

and taking inf (the largest upper bound) over all bijections, we get the Bottleneck distance between the diagrams (1):

$$W_\infty(X, Y) = \inf_{\eta : X \to Y} \sup_{x \in X} \|x - \eta(x)\|_\infty. \tag{8}$$

Wasserstein distance q-powers between X and Y:

$$W_q(X, Y) = \left[\inf_{\eta : X \to Y} \sum_{x \in X} \|x - \eta(x)\|_\infty^q \right]^{1/q}, \tag{9}$$

where $q > 0$.

Example 2. An example of free motion of a rigid body.

$$A\dot{\omega}_1 - (B - C)\omega_2\omega_3 = M_1,$$

$$B\dot{\omega}_2 - (C - A)\omega_1\omega_3 = M_2,$$

$$C\dot{\omega}_3 - (A - B)\omega_1\omega_2 = M_3.$$

Components of the angular velocity vector will be taken equal to zero: $M_1 = M_2 = M_3 = 0$.

a) Consider the case of a rigid body with the principal components of the inertia tensor: $A = 100.$, $B = 80.$, $C = 60$.

Initial values of the components of the angular velocity vector of a rigid body: $\omega_{x0} = \omega_{y0} = \omega_{z0} = 0.2$.

Step of integration of the system of differential equations: $h = 4$.

Barcodes of 3D images of the evolution of the components of the angular velocity vector of a rigid body.

Barcodes of dimension 0: 14[0.0, 0.05); [0.0, 0.1); [0.0, infinity).
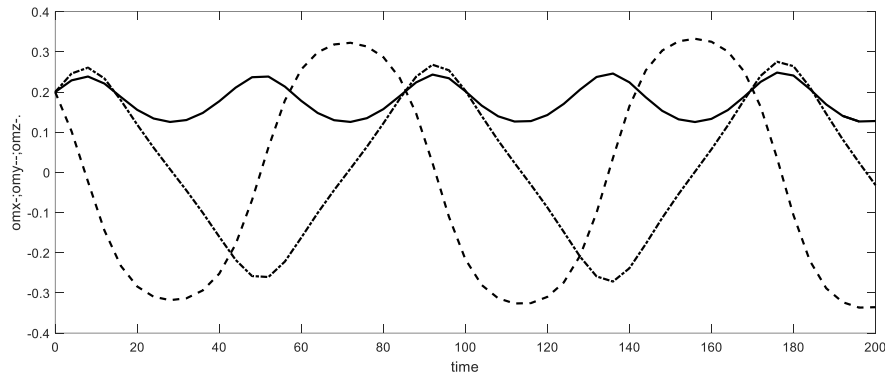
Barcodes of dimension 1: [0.1, 0.5).

Figure 1. Components of the angular velocity vector of a rigid body for the case of principal moments of inertia $A = 100.$, $B = 80.$, $C = 60.$: ω_x −; ω_x −−; ω_z −·.
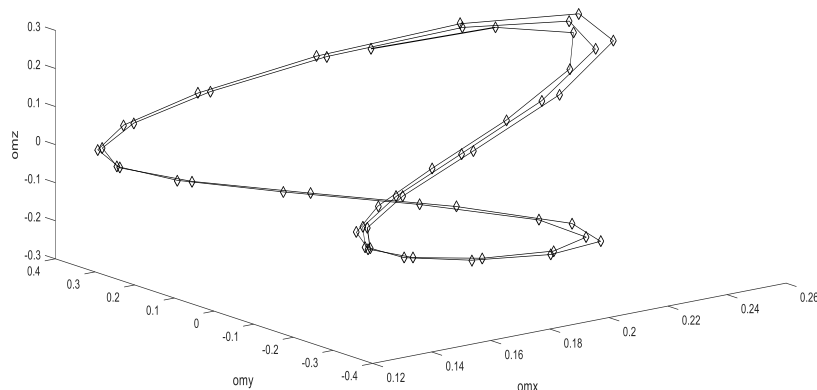

Figure 2. 3D image of the evolution of the components of the angular velocity vector of a rigid body for the case of principal moments of inertia $A = 100.$, $B = 80.$, $C = 60$.

b) Consider the case of a rigid body with the principal components of the inertia tensor: $A = 90.$, $B = 87.$, $C = 68$.

Initial values of the components of the angular velocity vector of the TT: $\omega_{x0} = \omega_{y0} = \omega_{z0} = 0.2$.

Step of integration of the system of differential equations: $h = 4$.

Barcodes of 3D images of the evolution of the components of the angular velocity vector of a rigid body.

Barcodes of dimension 0: 10[0.0, 0.05);

Barcodes of dimension 1: [0.05, 0.45).

Fig. 1 shows the graphs of the components of the angular velocity vector of a rigid body for the case of the principal moments of inertia $A = 100.$, $B = 80.$, $C = 60$.

Fig. 2 shows a 3D image of the evolution of the components of the angular velocity vector of a rigid body for the case of the principal moments of inertia $A = 100.$, $B = 80.$, $C = 60$.
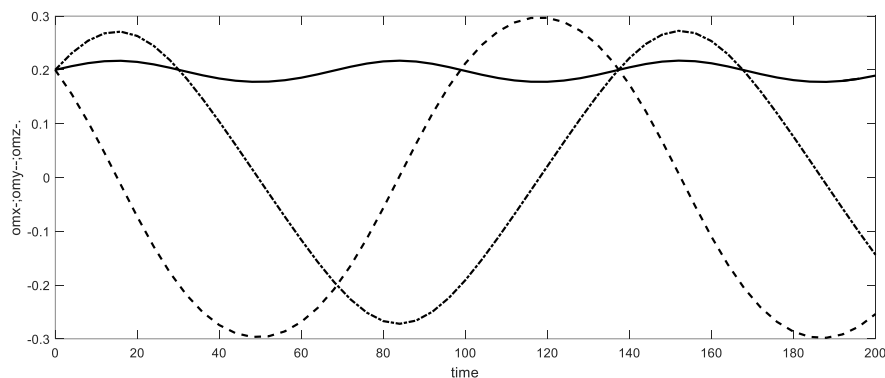

Figure 3. Components of the angular velocity vector of a rigid body for the case of principal moments of inertia $A = 90.$, $B = 87.$, $C = 68.: \omega_x \, -; \omega_x \, --; \omega_z \, -\cdot$.
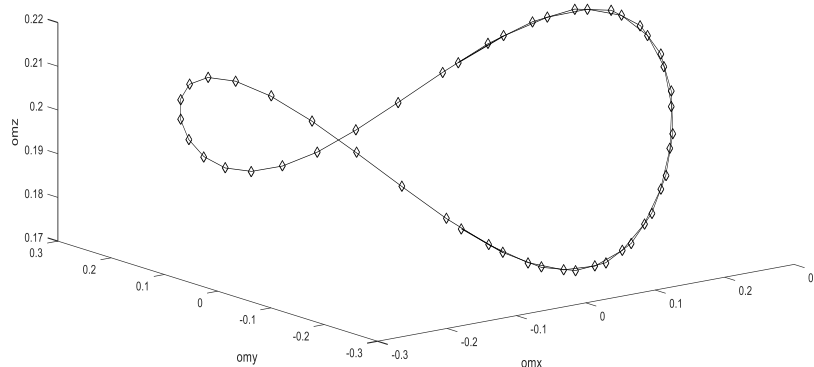

Figure 4. 3D image of the evolution of the components of the angular velocity vector of a rigid body for the case of principal moments of inertia $A = 90.$, $B = 87.$, $C = 68$.

Barcodes of 3D images of the evolution of the components of the angular velocity vector of a rigid body.

Barcodes of dimension 0: 10[0.0, 0.05);

Barcodes of dimension 1: [0.05, 0.45).

Fig. 3 shows the graphs of the components of the angular velocity vector of a rigid body for the case of the main moments of inertia $A = 90.$, $B = 87.$, $C = 68$.

Fig. 4 shows a 3D image of the evolution of the components of the angular velocity vector of a rigid body for the case of principal moments of inertia $A = 90.$, $B = 87.$, $C = 68$.

Bottleneck distance between 3D images (barcodes of dimension 1):

$$\|x - y\|_{\infty} = \max\{|x_1 - y_1|, |x_2 - y_2|\} = \max\{|0.1 - 0.05|, |0.5 - 0.45|\} = 0.05.$$

Example 3: Takens embedding for ω_y 3D $\tau(m-1) = 6$.

a) Consider the case of a rigid body with the principal components of the inertia tensor: $A = 100.$, $B = 80.$, $C = 60.$

Barcodes of dimension 1: $[0.15,\ 0.65)$.

Figure 5 shows a visualization of the 3D image of Takens embedding of the angular velocity component for the case of principal moments of inertia: $A = 100.$, $B = 80.$, $C = 60.$
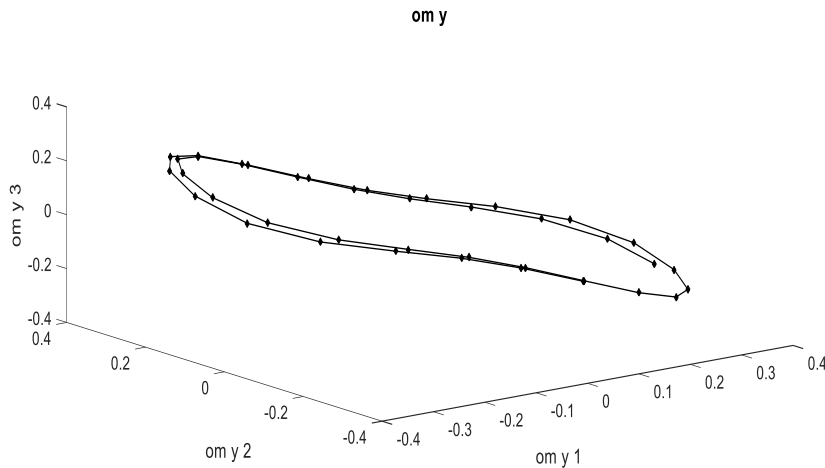

Figure 5. 3D image of the embedding of the Takens component ω_y for the case of principal moments of inertia $A = 100.$, $B = 80.$, $C = 60.$

b) Consider the case of a rigid body with the principal components of the inertia tensor: $A = 90.$, $B = 87.$, $C = 68.$

Barcodes of dimension 1: $[0.05,\ 0.6)$.

Figure 6 shows a visualization of the 3D image of Takens embedding of the angular velocity component for the case of principal moments of inertia: $A = 90.$, $B = 87.$, $C = 68.$
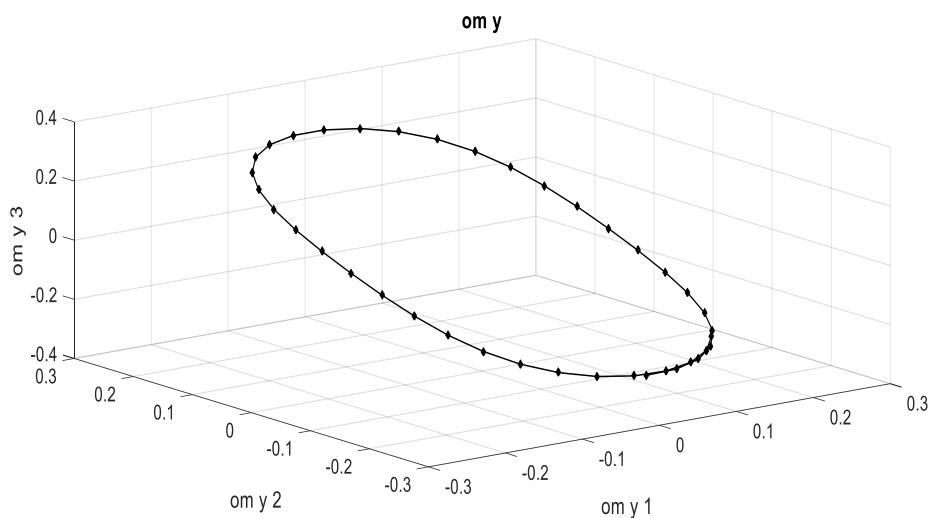

Figure 6. 3D image of the embedding of the Takens component ω_y for the case of principal moments of inertia $A = 90.$, $B = 87.$, $C = 68.$

Bottleneck distance between images (barcodes):

$$\|x - y\|_\infty = \max\left\{|x_1 - y_1|, |x_2 - y_2|\right\} = \max\left\{|0.15 - 0.05|, |0.65 - 0.6|\right\} = 0.1.$$

5. Conclusion

The paper considers the Takens embedding for two- and three-dimensional visualization of one-dimensional time series data.

The use of topological data analysis together with the Takens embedding to analyze the output information of a dynamic system - a rigid body is considered. 3D images of curves constructed from three components of the angular velocity vector of a rigid body for various values of the main moments of inertia are constructed. Three-dimensional images of the curves constructed on the basis of the Takens embedding for the component of the angular velocity of a rigid body for various values of the main moments of inertia are constructed.

Using the TDA method in conjunction with the Takens embedding for image comparison allows you to classify and identify images (or signals of a different physical nature).

The proposed TDA method, together with the Takens embedding, can be used for pattern recognition, data analysis in object control systems (for example, in an aircraft attitude control system).

The advantage of the TDA method lies in the invariance with respect to the Euclidean transformations of the components of the output information of devices and in the increase in the amount of analyzed information (in relation to traditional topological methods) due to the use of information about barcodes.

Acknowledgment

References

1. Edelsbrunner H., Harer J. L. Computational topology: an introduction. – American Mathematical Society, 2022.

2. Zomorodian A., Carlsson G. Computing persistent homology //Proceedings of the twentieth annual symposium on Computational geometry. – 2004. – pp. 347-356..

3. Carlsson G. Topology and data //Bulletin of the American Mathematical Society. – 2009. – Vol. 46. – №. 2. – pp. 255-308.

4. Wasserman L. Topological Data Analysis //arXiv e-prints. – 2016. – arXiv: 1609.08227.

5. Bourakna A. E. Y., Chung M. K., Ombao H. Topological Data Analysis for Multivariate Time Series Data //arXiv preprint arXiv:2204.13799. – 2022.

6. Pereira C. M. M., de Mello R. F. Persistent homology for time series and spatial data clustering //Expert Systems with Applications. – 2015. – Vol. 42. – №. 15-16. – pp. 6026-6038.

7. Gidea M., Katz Y. Topological Data Analysis of Financial Time Series: Landscapes of Crashes //arXiv preprint arXiv:1703.04385. – 2017..

8. Taken F. Dynamical systems and turbulence //Warwick, 1980. – 1981. – pp. 366-381.

9. Kennel M. B., Brown R., Abarbanel H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction //Physical review A. – 1992. – Vol. 45. – №. 6. – pp. 3403-3411.

10. Torku T. T. Takens Theorem with Singular Spectrum Analysis Applied to Noisy Time Series: Diss. – East Tennessee State University, 2016.

11. Kodba S., Perc M., Marhl M. Detecting chaos from a time series //European journal of physics. – 2004. – Vol. 26. – №. 1. – pp. 205.

12. Offroy M., Duponchel L. Topological data analysis: A promising big data exploration tool in biology, analytical chemistry and physical chemistry //Analytica chimica acta. – 2016. – Vol. 910. – pp. 1-11.

13. Perea J. A., Harer J. Sliding windows and persistence: An application of topological methods to signal analysis //Foundations of Computational Mathematics. – 2015. – Vol. 15. – №. 3. – pp. 799-838.

14. Maletić S., Zhao Y., Rajković M. Persistent topological features of dynamical systems //Chaos: An Interdisciplinary Journal of Nonlinear Science. – 2016. – Vol. 26. – №. 5. – pp. 053105.

15. Chukanov S.N. Comparison of object images by methods of computational topology. // Informatics and automation. – 2019. – Vol. 18. – №. 5. – pp. 1043-1065.

16. Leikhter S.V., Chukanov S.N. Comparison of images based on their diffeomorphic transformation. // Computer Optics. – 2018. – Vol. 42. – № 1. – pp. 96-104.

17. Hatcher A. Algebraic topology. – Cambridge UP. – 2005.

18. Bubenik P. The persistence landscape and some of its properties // Topological Data Analysis. – Springer, Cham, 2020. – pp. 97-117.