Topological data analysis of Lagrangian orbits in natural convection flows confined in a cylinder

José Núñez *
Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autónoma de México, 58190 Morelia, Michoacán, México

Ahtziri González
Facultad de Ingeniería Eléctrica, Universidad Michoacana de San Nicolás de Hidalgo, 58030 Morelia, Michoacán, México

Eduardo Ramos
Instituto de Energías Renovables, Universidad Nacional Autónoma de México, 62580 Temixco, Morelos, México

(Received 17 January 2022; accepted 8 November 2022; published 9 December 2022)

This article presents a topological data analysis of the Lagrangian orbits occurring in natural convection flows inside cylindrical containers heated from below. The fluid motions are calculated via the numerical integration of the mass, momentum, and energy conservation equations using the Fourier-Chebyshev pseudospectral method. The study focuses on steady-state flows, and their qualitative properties are described in terms of the Rayleigh number. The Lagrangian orbits are calculated using the Eulerian velocity fields and their features are analyzed with Poincaré maps placed at the horizontal central plane of the cylinder. The qualitative features of the Lagrangian orbits calculated are closely related to those predicted by Hamiltonian theory, with the phase space being the real space in our case and considering the Rayleigh number as the driving parameter. The orbits are embedded in nested tori surfaces for low Rayleigh numbers ($\mathrm{Ra} = 8 \times 10^3$). Some orbits seem to be dense in the tori and others form sets of points divided into segments in the Poincaré maps. At larger Rayleigh numbers ($\mathrm{Ra} = 10^4$, 2×10^5) the orbits form increasingly complicated structures, including loop islands in the Poincaré maps. Also, large areas in the maps show irregular or chaotic distributions of points. Topological data analysis (TDA) was used to define parameters that quantify the geometrical properties of the points distributions in the Poincaré maps. The number of segments in the sets of points and the occurrence of holes were determined with the 0- and 1-persistent homologies, respectively. Pairs of relatively prime numbers are used to characterize quasiperiodic orbits; the pairs are naturally identified with torus knots. It is found that the ratio of toroidal to the poloidal number of turns follows a simple rule for each Rayleigh number explored. The most important advantage of using TDA is the quantification and accurate identification of topological features found in the Poincaré maps.

DOI: 10.1103/PhysRevFluids.7.123501

I. INTRODUCTION

Natural convection in containers has been profusely analyzed due to its importance in many engineering applications like crystal growth and liquid metal batteries. This flow has also been

*joseng@enesmorelia.unam.mx

looked into as a simple example of complex dynamics due to its inherent nonlinear nature. The phenomenon can be conceived as a Rayleigh-Bénard type problem where the walls of the container have a definite influence. Informative and comprehensive reports on the properties of the flow are now available in the literature [1–4].

Qualitative properties of natural convection in vertical cylinders heated from below as functions of the Rayleigh (Ra) and Prandtl (Pr) numbers and aspect ratio ($A =$ height/diameter) are now well established through experimental and numerical simulation. See, for instance, the pioneering work reported in Refs. [5] and [6]. For an aspect ratio of 1.25 and a Prandtl number of 6.67, which corresponds to the conditions of interest in this report, the fluid remains motionless for $\mathrm{Ra} \lesssim 5 \times 10^3$. A steady motion is detected for intermediate Rayleigh numbers $5 \times 10^3 \lesssim \mathrm{Ra} \lesssim 10^6$, and a time-dependent flow is found for larger Rayleigh numbers [5]. A detailed numerical and experimental study of the natural convective flow of water in a cylindrical container. with an aspect ratio of approximately 1.3 was presented by Núñez *et al.* [7,8]. In a steady state, the flow is dominated by a large nonaxisymmetric vortex with a horizontal axis. In a specific vertical plane of symmetry that contains the cylinder's symmetry axis, it is found that the azimuthal velocity is zero. This plane effectively divides the volume into two halves with the same dynamics.

Most studies of natural convection in containers have been made using experimental and numerical Eulerian tools, where velocity and other variables are defined as functions of space and time. Lagrangian analyses are more scarce, even though they are relevant for understanding mixing and passive and nonpassive scalar transport in flows [9]. The Lagrangian orbits form structures that have been previously described in two-dimensional (2D) time-dependent and 3D steady flows and that have been interpreted in terms of Hamiltonian mechanics and symplectic maps [10]. The structures include nested tori that become unstable at a critical number of the forcing parameter to generate Kolmogorov-Arnold-Moser (KAM) tori, island chains around KAM tori (resonance zones) and chaotic regions around tori and islands [11,12]. The general theory that links Hamiltonian mechanics and properties of Lagrangian orbits is known as chaotic advection and has seen enormous development since the early 2000s. The Lagrangian study of convective flows has greatly benefited from the previously described identification [10].

A Lagrangian analysis of three-dimensional steady natural convective flow in a cubic container with the Grashof number (Gr) as the driving parameter that is closely related to the present study has recently been published [13]. For small Gr, the Lagrangian orbits form nested toroidal structures whose Poincaré maps in the midhorizontal plane show points distributions similar to those found in the 2D maps of the area preserving Hamiltonian systems. Increasing Gr causes the progressive disintegration of tori into chaotic streamlines following universal Hamiltonian mechanisms governed by the KAM and Poincaré-Birkhoff theorems [14,15]. It is interesting to observe that a mechanically generated flow (double lid-driven flow) displays an equivalent topological behavior [16]. Additionally, it is worth noticing that the orbits of electrically charged particles share some geometrical properties with the Lagrangian orbits reported in the natural convective flows [17].

In the present study, topological tools were applied to analyze the geometric features of the Poincaré map defined by the Lagrangian orbits. We integrate the conservation equations with a spectral method to obtain a highly accurate solution. This is required to obtain reliable results in the analysis of the Poincaré maps. This point was remarked on by Ravu *et al.* [18], who showed how the high accuracy of the numerical method gives qualitatively and quantitatively superior trajectories that result in more accurate identification of Lagrangian structures. Once the velocity field is calculated, particle tracking is performed, in which a high-order approximation is implemented to achieve good precision of trajectories at large times. The points clusters in the Poincaré map are studied with topological data analysis (TDA). In databases analysis, it is well known that the shape of data points provides important information on the behavior of the studied phenomenon [19,20]. Currently, TDA has been recognized as a very useful tool for obtaining meaningful information from sets of points, because it is computationally light and robust to small variations in the data. This technique has been used to analyze data clouds obtained from a great variety of phenomena, including climatic behavior [21], chaos detection in time series [22,23], and in other areas [24].

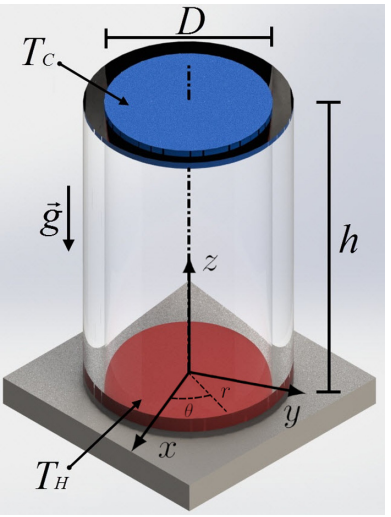

FIG. 1. Geometry of the cylindrical container.

The idea of studying the properties of flows with topological methods is not new. Early applications of topology tools to fluid mechanics focused on the study of properties and evolution of vortex filaments and linked structures interpreted as knots [25,26]. More recently, other topological tools have been used to expand the scope of applications. A description of tools that have been used in the analysis of flows can be found in monographs and review papers, for instance, Ref. [27]. Specifically, persistent homology has been used to characterize and detect bifurcations and chaos in complex systems. An important example of the application of persistent homology concepts is the study by Kramár *et al.* [28], they describe the temperature field in Rayleigh-Bénard convection in a shallow cylindrical container with a small aspect ratio. They concluded that this tool is a robust and effective method to study periodic dynamics and identify equilibria. A similar study to the present investigation is reported in Ref. [23]; the authors claim that persistent homologies can be used for the early detection of bifurcations.

The rest of this paper is organized as follows: Section II describes the physical problem, the mathematical model, and the numerical methodology. Section III contains examples of applications of topological data analysis to point clouds in the plane. Section IV describes the main results, and, finally, Sec. V draws the main conclusions of this work. Given that the persistent homologies have not been used frequently in flow analysis, we include relevant information in the Appendix and in Supplemental Material [29]. In these documents, one can find a general and intuitive introduction to TDA for approximating data clouds with complexes and extracting information by using persistent homologies. Readers familiar with persistent homology concepts may wish to skip the Appendix and Supplemental Material [29].

II. PHYSICAL SYSTEM, THEORETICAL MODEL, AND NUMERICAL STRATEGIES

A. Problem formulation

We consider a circular cylindrical container with a height h and diameter D, the aspect ratio of the cylinder is $h/D = 1.25$. The computations were made in cylindrical coordinates, with the origin at the center of the bottom cap. The axis of symmetry of the cylinder corresponds to the z axis (Fig. 1).

The lower wall of the cylinder is kept at a hot temperature T_H, and the upper wall is kept at a low temperature of T_C. The lateral wall is thermally insulated, and the working fluid is an incompressible

Newtonian fluid with $\mathrm{Pr} = 6.67$ with constant kinematic viscosity ν, thermal diffusivity α, and volume expansion coefficient β.

B. Mathematical model

The transport equations needed to describe the phenomenon are the conservation of mass, momentum, and energy. We assume that the temperature difference $\Delta T = T_H - T_C$ is small and that the density variation is only present in the buoyancy term (Boussinesq approximation).

This set of equations has been solved in its dimensionless form. The scaling we used was the most convenient for the numerical solution and includes the following characteristic quantities: The axial and radial coordinates are expressed in terms of the height (h) of the cylinder, respectively. The characteristic velocity (u_c) is the free-fall velocity, namely $u_c = \sqrt{g\beta\Delta T h}$, where g is the terrestrial gravity acceleration. The timescale is defined using the characteristic velocity as h/u_c. In dimensionless form, the steady governing equations are written in cylindrical coordinates (r, θ, z) as follows:

$$\frac{1}{r}\frac{\partial}{\partial r}(ru_r) + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} = 0, \tag{1}$$

$$(\vec{u}\cdot\nabla)u_r - \frac{u_\theta^2}{r} = -\frac{\partial p}{\partial r} + \left(\frac{\mathrm{Pr}}{\mathrm{Ra}}\right)^{\frac{1}{2}}\left(\nabla^2 u_r - \frac{u_r}{r^2} - \frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta}\right), \tag{2}$$

$$(\vec{u}\cdot\nabla)u_\theta + \frac{u_\theta u_r}{r} = -\frac{1}{r}\frac{\partial p}{\partial \theta} + \left(\frac{\mathrm{Pr}}{\mathrm{Ra}}\right)^{\frac{1}{2}}\left(\nabla^2 u_\theta - \frac{u_\theta}{r^2} + \frac{2}{r^2}\frac{\partial u_r}{\partial \theta}\right), \tag{3}$$

$$(\vec{u}\cdot\nabla)u_z = -\frac{\partial p}{\partial z} + \left(\frac{\mathrm{Pr}}{\mathrm{Ra}}\right)^{\frac{1}{2}}\nabla^2 u_z + T, \tag{4}$$

$$(\vec{u}\cdot\nabla)T = \left(\frac{1}{\mathrm{RaPr}}\right)^{\frac{1}{2}}\nabla^2 T, \tag{5}$$

where $\vec{u} = (u_r, u_\theta, u_z)$ is the velocity, p is the pressure, and T is the temperature. The differential operators are

$$\vec{u}\cdot\nabla = u_r\frac{\partial}{\partial r} + \frac{u_\theta}{r}\frac{\partial}{\partial \theta} + u_z\frac{\partial}{\partial z}$$

and

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}.$$

The flow is characterized by two dimensionless parameters, the Rayleigh and Prandtl numbers, defined by

$$\mathrm{Ra} = \frac{g\beta\Delta T h^3}{\nu\alpha} \quad \text{and} \quad \mathrm{Pr} = \frac{\nu}{\alpha}. \tag{6}$$

The boundary conditions corresponding to the physical situation of interest are hot and cold walls

$$T = 1 \quad \text{on} \quad z = 0, \quad T = 0 \quad \text{on} \quad z = 1, \tag{7}$$

the insulating boundary at the sidewall

$$\frac{\partial T}{\partial r} = 0 \quad \text{on} \quad r = D/2h \tag{8}$$

and

$$\vec{u} = 0 \quad \text{in all boundaries.} \tag{9}$$

C. Numerical solution

The numerical integration of the conservation equations is accomplished with a spectral method [30]. The governing equations were discretized using the Fourier-Galerkin method in the azimuthal direction and the Chebyshev pseudospectral (or collocation) method for the radial and axial directions [31]. Applying spectral methods, high-order approximation of the partial derivatives is obtained, and no staggered grids are used. Using the Fourier method, the solution naturally satisfies the azimuthal periodic boundary conditions. The time derivative is approximated using the second-order Adams-Bashforth method. Matrix derivatives were calculated with polynomial interpolation formulas. The explicit nonlinear terms are computed in the physical space and carried into the Fourier space through a discrete transform. A projection method was used as the pressure-velocity decoupling strategy. The resulting linear system of equations is solved fast and efficiently by a diagonalization method. As it will be commented further in the next subsection, the satisfaction of mass conservation by the velocity field is of particular importance for reliable tracking because this ensures accurate preservation of the Hamiltonian nature of the Lagrangian trajectories. In the numerical solution of the convective flow reported here, a velocity divergence error smaller than 10^{-10} was obtained.

As a validation, the numerical solutions are compared to experimental results in Refs. [7,8]. It was found that the flow simulations quantitatively predict the experiments, and although the comparison was made using the Eulerian description, e.g., three-dimensional steady velocity fields, we consider this is strong evidence to suggest that the numerical solutions are correct.

D. Lagrangian tracking

The structure of the flow is visualized with Lagrangian trajectories, which are the lines formed by subsequent positions of massless, nondiffusive particles whose velocity coincides with that of the fluid at a particular point. The Lagrangian trajectories are calculated in the steady state, assuming that the particle velocity is constant according to the following expressions:

$$u_\theta = u_\theta(r, \theta, z), \quad u_r = u_r(r, \theta, z), \quad u_z = u_z(r, \theta, z). \tag{10}$$

The rate of change of the position vector is

$$\dot{r} = u_r, \quad r\dot{\theta} = u_\theta, \quad \dot{z} = u_z. \tag{11}$$

A fourth-order multidimensional Runge-Kutta method (RK4) was used to solve the problem. The definition of the initial time t_0 is arbitrary, but as described in detail below, the geometrical properties of the orbits are dependent on the point $\vec{x}_0$ where the integration of Eq. (11) starts (initial condition). The accuracy of Lagrangian tracking is of crucial importance because errors can compromise the Hamiltonian properties of the orbits. See Refs. [13,18]. In this study, we have paid special attention to this point. Besides the accurate mass conservation reflected by the small error in the velocity divergence commented on in the previous subsection, we have used precise interpolation and an RK4 temporal integration method. The spatial interpolation method is of order $nr \times nz$, where n is the number of points.

A frequently used tool to describe the geometrical properties of the orbits is the Poincaré maps, which can be built by choosing a plane and recording the positions where the orbit crosses the plane. Formally, Poincaré maps can be regarded as maps of a plane on itself [32]. The Poincaré maps are built with the intercepts of the orbits with the horizontal plane at $z = 0.5$. To get a full picture of the dynamics in the volume, it is necessary to use several initial conditions. We considered 22 initial points to cover a wide region in the plane.

Topological properties of the Lagrangian trajectories represented in the Poincaré map were analyzed with TDA. A comprehensive guide of topological data analysis can be found in Refs. [33,34]. To make this document self-contained, we provide a brief introduction to TDA in Appendix.

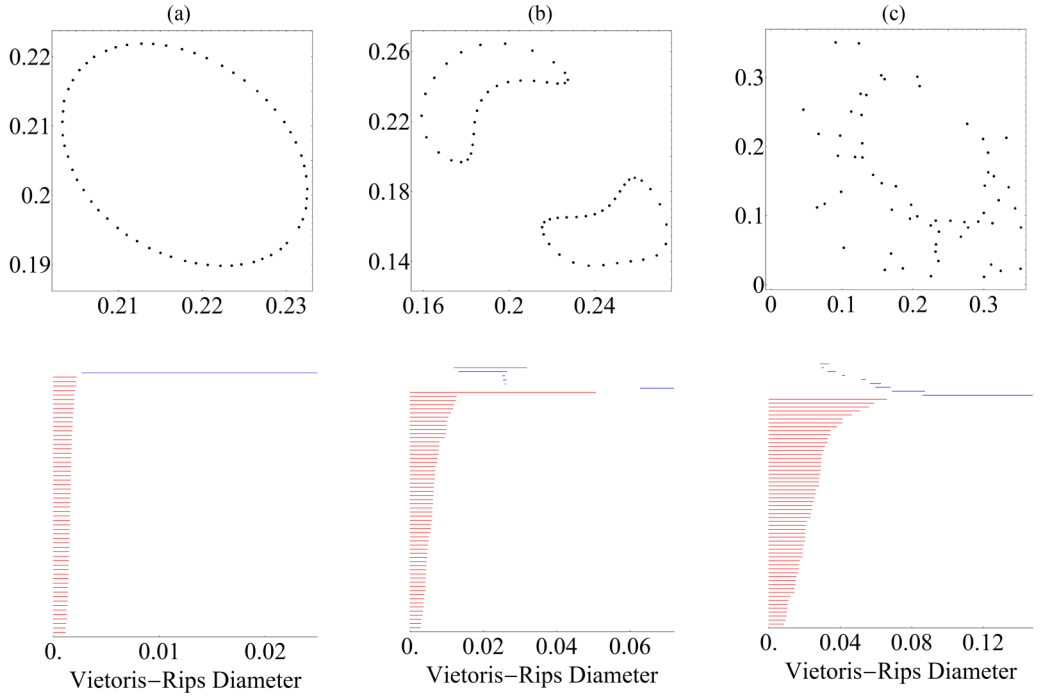

FIG. 2. Examples of Vietoris-Rips barcodes for three data clouds in $\mathbb{R}^2$. Bars for 0-homologies and 1-homologies appear in red and blue, respectively.

III. EXAMPLES OF TOPOLOGICAL DATA ANALYSIS APPLICATIONS

In this section, we describe three examples of TDA to introduce the details of the methodology. The point distributions chosen as examples are closely related to the Poincaré maps formed by the Lagrangian orbits in the convective flow that are described in Sec. IV.

We compute the persistent homology with the TDA package included in the R program [35]. In the output, the barcodes are shown as follows: The 0-homologies (red bars) start at the smallest radius for the Vietoris-Rips complex considered and stop when the Vietoris-Rips complex of the corresponding radius is connected. See the Appendix for an intuitive explanation of Vietoris-Rips complexes, homology, persistent homology, and barcodes. The 1-homologies (blue bars) start at the radius when the first hole is formed and end when there are no holes in the Vietoris-Rips complex. Individual bars in the 0-homology are displayed piled up according to the ascending order in the filtration radius. In the 1-homology, the bars are arranged such that the one that is born first occupies the uppermost position and subsequent ones are drawn under the previous bar.

Three examples of barcodes are shown in Fig. 2. These examples were chosen for their similarity with the point clusters found in the Poincaré maps of Sec. IV. We proceed to describe the information contained in these examples. Our goal is to illustrate that many geometrical properties of point clusters can be quantitatively described by the barcodes. The properties described in the following paragraphs can be more fully understood by looking at the barcodes formation as the Vietoris-Rips radii grow. This information is in the videos in Supplemental Material [29].

In the first example that corresponds to Fig. 2(a), all red bars (0-homologies) are short and display small length increments, indicating that the connectivity of the set of points is also increased by small amounts as the radii grow. An equivalent interpretation is a small increment in radii results in one or a small number of new 1-simplexes (segments). From this property, we deduce that the distances between pairs of points are similar, and the points are evenly distributed in the cloud.

In contrast, there is only one blue bar of a relatively larger length, indicating the presence and long persistence of a single hole. The line is born at radii slightly larger than the longest red bar. Combining the two features revealed by the homologies, we infer that the points are arranged in a loop configuration.

In the second example [Fig. 2(b)], the red bars display small increments up to the one before the last, where a substantial length increase is observed. The interpretation of this feature is that in most of the exploration, the connectivity of the points set increases slowly, as in Fig. 2(a). The substantial size difference between the two uppermost bars indicates that the set is composed of two separate clusters, and it is only when large enough radii permit the appearance of a segment joining the two clusters then the last bar dies. The blue bars (1-homology) refer to the formation of 1-holes. We find three long bars that indicate dominant persistent features of the points distribution. The two of them that exist for approximately the same range of radii correspond to the holes inside each of the kidneylike loops. The blue bar on the right corresponds to the hole that is formed between the loops when segments join points from the two separate clusters. Note that none of the holes persist for long radii intervals because they fill up as the radii grow. On top of the three bars previously described, there are also three tiny blue bars that reveal the formation of short-lived 1-holes. These do not represent persistent features of the point distribution. The remarks given in this paragraph are further illustrated in Supplemental Material [29].

In the third example, shown in Figure 2(c), consecutive red bars do not show a big increase in length. Also, the longest red bar indicates that we need a considerably big radius to connect all the points in the set (close to half the radius necessary for the Vietoris-Rips complex to be the convex hull). This means that the points are scattered in the data cloud but do not form clusters. The appearance of small holes with little persistence that are formed and filled after a small radii increase can be deduced from the shorter blue bars. The longest-lived hole appears at relatively large radii and shows persistence, indicating that the dots form a large hollowed structure similar to the one described in Fig. 2(a). The five intermediate-sized blue bars plus the largest indicate that the points cloud has five small looplike structures arranged around the more persistent hole. It is interesting to remark that the length of the longest red bar is approximately equal to half the radius at which the last blue bar dies. This confirms our previous observation that the points are scattered and do not form definite clusters.

IV. RESULTS

A. General properties of the flow

The results presented in this section were obtained by analyzing the dynamics of a natural convection flow of a Newtonian fluid with a Prandtl number $\mathrm{Pr} = 6.67$ confined in a cylindrical container with an aspect ratio of 1.25. We study the Lagrangian orbits in flows with Rayleigh numbers $\mathrm{Ra} = 8 \times 10^3$, 8×10^4, 2×10^5, and 5×10^5. In all cases studied, the fluid is in a steady state. As is well established, under the conditions of interest, the fluid is motionless for Rayleigh numbers less than $\mathrm{Ra}^{c_1} \approx 7 \times 10^3$ and acquires a time-dependent motion for Rayleigh numbers greater than approximately $\mathrm{Ra}^{c_2} \approx 10^6$ [5]. In the four cases, the flow is symmetric with respect to a vertical plane. For this reason, we will only describe the properties in half the volume.

The Lagrangian orbits are defined by mass-less particles moving through the fluid with a local, instantaneous velocity dictated by the Eulerian velocity field obtained from the numerical solution of Eqs. (1)–(5). The structure of the Lagrangian orbits in a three-dimensional steady flow in a closed domain has been identified with Hamiltonian dynamics, and a large body of knowledge on the subject is now available, see, for instance, Refs. [9–11]. As expected, we have also found the generic properties explained in the literature, but to give a background to the quantitative analysis proposed in this study, we include in this section a qualitative description of the structures of the Lagrangian orbits found in the natural convection flow in a cylindrical container.

Figure 3 shows some examples of Lagrangian orbits for each Rayleigh number studied. Notice that some of them seem to move along the surfaces of tori. The Lagrangian orbits naturally

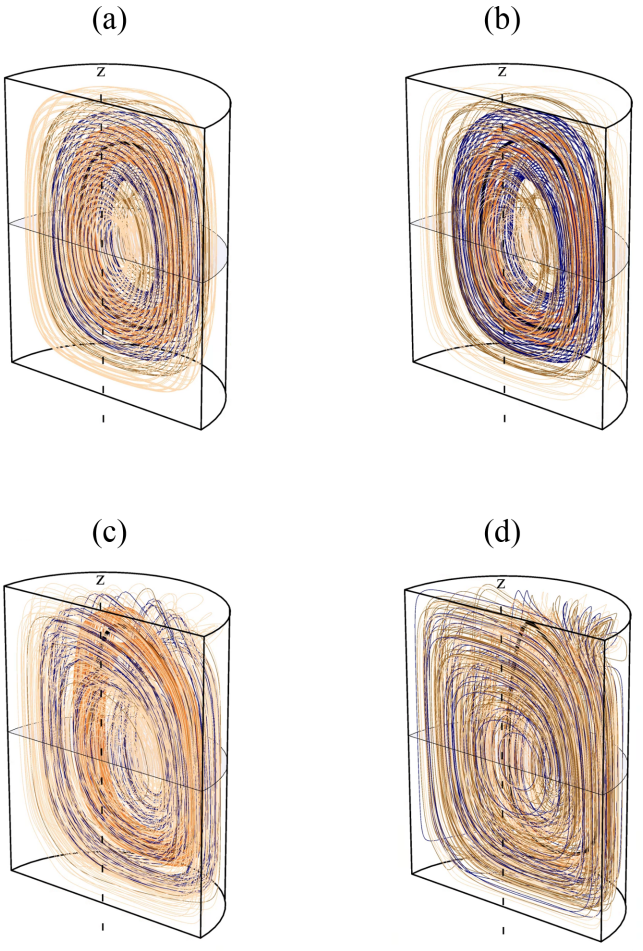

FIG. 3. Lagrangian orbits determined by the flow in one half of the cylinder. Each image shows only five colored orbits in order to appreciate the internal structure of the flow. (a) $\mathrm{Ra} = 8 \times 10^3$, (b) $\mathrm{Ra} = 10^4$, (c) $\mathrm{Ra} = 2 \times 10^5$, and (d) $\mathrm{Ra} = 5 \times 10^5$.

divide the studied region into two parts, on one side the orbits go up and on the other, they go down.

To study the flow in detail, we use the Poincaré map in the plane perpendicular to the cylinder axis at midheight. The Poincaré map is a widely used tool to study flows since it allows us to do the analysis in a dimension one smaller and preserves the properties of the Lagrangian orbits [36]. In our specific case, the map consists of the collection of points in the trajectories that belong to the horizontal plane at height $z = 0.5$. Lagrangian orbits are identified with different colors and each orbit defines a set of points with the same color in the Poincaré map.

Figure 4 shows the Poincaré maps of the four Rayleigh numbers studied. In each case, the map is divided into two sides separated by a vertical line. Although the dot distributions in the two regions are not entirely symmetric, we only analyze the properties of the sets of points on the right side because each point on the left side is connected to a point on the right side through the orbit which, naturally defines a bijective function.

In all four cases, the points defined by a Lagrangian orbit in the Poincaré map surround a center (black points in the images of Fig. 4). The centers were determined by experimentation with a

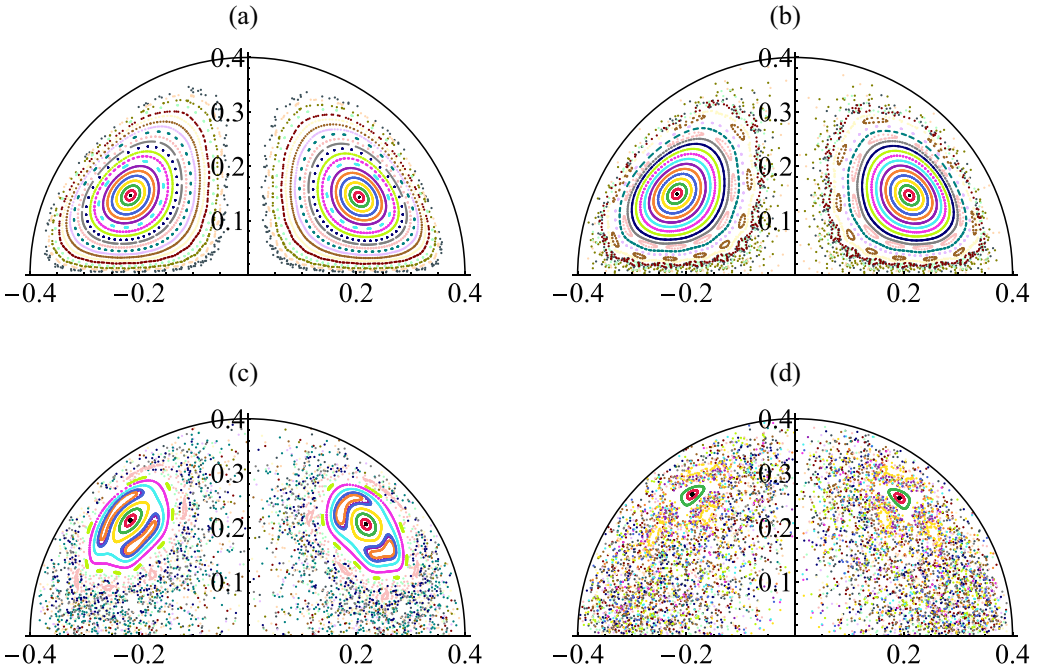

FIG. 4. Poincaré maps at the horizontal middle plane of the cylinders. The color of the points corresponds to each of the 22 Lagrangian orbits studied. Note that the black dots are the central trajectories. (a) $\mathrm{Ra} = 8 \times 10^3$, (b) $\mathrm{Ra} = 10^4$, (c) $\mathrm{Ra} = 2 \times 10^5$, and (d) $\mathrm{Ra} = 5 \times 10^5$.

precision of 10^{-3} and their positions in each Rayleigh number studied are $c_1 = (0.205371, 0.143111)$, $c_2 = (0.21014, 0.146272)$, $c_3 = (0.217764, 0.205967)$, and $c_4 = (0.192143, 0.254469)$.

For the analysis, we took 22 Lagrangian orbits for each Rayleigh number, and the corresponding starting points were chosen equidistant from the center (including c_i) to the vertical radius of the cylinder. The Poincaré maps in Fig. 4 were obtained from Lagrangian orbits with these initial conditions.

The centers of the Poincaré maps determine closed Lagrangian orbits (see the internal black orbit in each of the four cases shown in Fig. 3). Similarly, the sets of points that seem completely circular in the Poincaré map, determine stream surfaces in the cylinders with the shape of a torus. The corresponding orbit lives on such a stream surface.

The features observed in the Poincaré maps of the Lagrangian orbits shown in Fig. 4 are closely related to the points distributions obtained with area preserving Hamiltonian systems. This is a result of the Hamiltonian structure of the equations of motion.

A summary of the map's features can be described as follows:

(1) $\mathrm{Ra} = 8 \times 10^3$. The map is composed of sets of points arranged in loops nested around the common center c_1. This is a one-period fixed point because the Lagrangian orbit always returns to it on the Poincaré map after one turn. In some loops, the points appear to be uniformly distributed, while in others, the points are grouped forming linear islands (short line segments).

(2) $\mathrm{Ra} = 10^4$. The 12 innermost sets of points form nested loops, some of them consisting of linear islands around the fixed point c_2. The five middle sets are formed by islands; in two of them, the islands are organized in clearly defined loops. In the context of dynamical systems, these structures are known as elliptic and hyperbolic critical point chains. The outermost four sets display less ordered, scattered points located in the outer part of the region under analysis.

(3) $\mathrm{Ra} = 2 \times 10^5$. The four sets of points with a smaller average distance to the center (this parameter is denoted $\bar{d}$ and is formally defined in the next section) are nested loops arranged around the point c_3 in a similar setting to those observed for smaller Rayleigh numbers. The three clusters located further away from c_3 are divided into two separate looplike structures. The middle three sets of points define loops (one of them with linear islands) that surround the inner structures. The 11 exterior sets are formed by scattered points. Observe, however, that one of them forms not-well-defined loop islands.

(4) $\mathrm{Ra} = 5 \times 10^5$. The first three sets form nested loops around the center c_4. The remaining set of points is randomly distributed and fills the rest of the region. There is one exception where the set looks organized on closed islands (yellow points).

The point distributions are similar to those found in the analysis of simplectic maps (e.g., the standard map), with the Rayleigh number being the driving parameter [11,14,37]. In terms of the nomenclature of Hamiltonian dynamics, we identify the sets of points as follows. The loops are KAM tori. The sets divided in linear islands correspond to the chains around KAM tori that result from their break-up due to resonance. Our sets divided into loop islands are called island chains. The scattered point sets are the points in the chaotic regions around tori and islands.

In the next subsection, we will quantify the geometrical properties of the distribution of points observed in the Poincaré maps using topological tools, which allow us to uncover properties that would have been difficult or impossible to discover with the qualitative methodology sketched in the previous paragraphs. Readers not familiar with TDA can find a brief and intuitive introduction in the Appendix. See Supplemental Material [29] for examples and illustrations of the processes used to get the topological properties of the sets of points.

B. Topological data analysis

In this section, we use TDA to study the global structure of the Poincaré maps in Fig. 4. We divide the Poincaré maps into sets of points corresponding to one individual Lagrangian trajectory; then we compute the 0- and 1-persistent homologies for each one of these sets. Given a set of points X in the Poincaré map defined by one of the Lagrangian trajectories, the parameters that we use to describe the results are the following:

(1) The average distance $\overline{d}$ of points in X to the center c_i (see Fig. 4).

(2) The length of the longest bar in the 0-persistent homology of X barcode, VR_0. Observe that VR_0 is the first radius for which the Vietoris-Rips complex of X is connected.

(3) The final bar in the barcode of the 1-persistent homology of X. This bar dies at the first radius for which the Vietoris-Rips complexes do not have holes.

(4) The number of clusters in which X is divided will be denoted p. Note that as explained in Fig. 2(b), $p-1$ is the number of significant jumps in the lengths of two contiguous bars in the barcode of the 0-persistent homology of X.

(5) The number of bars in the barcode of the 1-persistent homology of X will be labeled $\mathcal{N}$. This parameter indicates the number of holes detected.

(6) The sum of all but the last 1-persistent homology bars of X is denoted $\mathcal{E}$. A simple distribution of points like those in Fig. 2(a), will give $\mathcal{E} = 0$, while increasingly more complex distributions like Figs. 2(b) and 2(c) in the same figure, give $\mathcal{E} = 0.034401197$ and $\mathcal{E} = 0.045860959$, respectively.

The information for each of the four Rayleigh numbers analyzed is presented in two separate bar graphs, one for 0- and another for 1-homology. The bars in the following figure do not correspond to the barcodes. The color of the bars corresponds to the color of the Lagrangian orbits in the Poincaré map of Fig. 4. The bars are arranged in ascending order according to the choice of the initial points of the orbits, and the average distance $\overline{d}$ is displayed on the left-hand side of each bar.

In the 0-homology graph, each bar is the longest in the 0-persistent homology barcode of a set of points defined by a Lagrangian trajectory (i.e., the length of a bar, VR_0). The number of components that are separated by significant jumps (p) is displayed on the right-hand side of each bar.

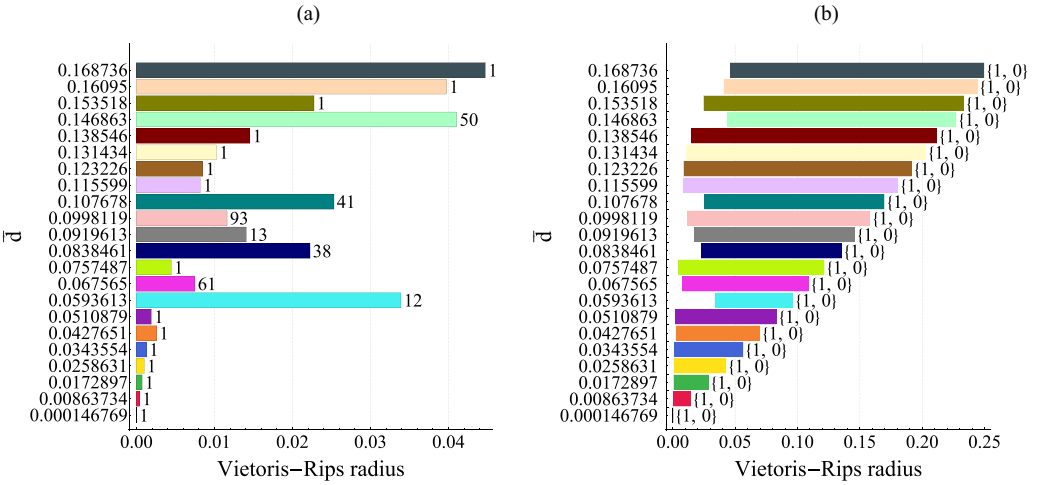

FIG. 5. TDA for $\mathrm{Ra} = 8 \times 10^3$. The bars correspond to each of the 22 Poincaré sets of points of the Lagrangian orbits studied. See Fig. 4(a). The color of the bars matches the color of the sets of points. The average distance $\overline{d}$ is shown in the column on the left-hand side. The left panel shows the longest bars (VR_0) of the 0-homology analysis and the number of components in each set of points (p) is indicated at the end of each bar. The right panel shows the last bars for the 1-homology analysis and the parameters $\{\mathcal{N}, \mathcal{E}\}$ are given on the right-hand side of each bar.

In the 1-homology graph, each bar corresponds to the last bar in the 1-persistent homology barcode of a point set defined by a Lagrangian trajectory. The number of bars in the 1-persistent homology barcode and the sum of the length of all but the last one are displayed as $\{\mathcal{N}, \mathcal{E}\}$ on the right-hand side of each bar. It is important to emphasize that $\mathcal{N}$ represents the total number of 1-holes, not only the persistent ones; refer to Figs 2(b) and 2(c).

In the following paragraphs, we proceed to spell out the information contained in the two graphs of the four cases studied. To simplify the description, we start from the lower bars to the upper ones.

Figure 5, $\mathrm{Ra} = 8 \times 10^3$, the graph on the left, which corresponds to 0-homology, indicates that the average distance $\bar{d}$ increases monotonically. The graph on the right shows that $\mathcal{N} = 1$ and $\mathcal{E} = 0$ for all sets of points, then they have just one hole, and the points are distributed along simple loops. Also, we observe that in the 1-homology graph, the Vietoris-Rips radii where the bars die, i.e., where the final 1-hole is filled, is a monotonously increasing function of $\bar{d}$. Assembling the information from the two graphs, we conclude that the sets of points do not mix with each other. Consequently, they form nested loops with increasing average distances around the center c_1. The bars show that seven sets with smaller $\overline{d}$ display similar properties, with evenly separated points forming thin loops. All sets of points but one in the next seven sets are grouped in clearly separated clusters (linear islands) as indicated by $p > 1$ and the bar lengths (VR_0). The previous information is consistent with the fact that the flow is divided into stream surfaces which are nested tori, as suggested in Fig. 3(a).

The left graph of Figure 6, $\mathrm{Ra} = 10^4$ indicates that the average distance $\overline{d}$ of the first 18 sets of points increases monotonously, and similarly to the previous case, from the two graphs, we conclude that the first 11 sets of points form nested simple loops around c_2. Some of these sets are divided into linear islands, as indicated by the parameter $p > 1$ on the 0-homology graph. Also, the graphs show that $\overline{d} = 0.101278$ and $\overline{d} = 0.129062$ sets of points (salmon and brown) are divided respectively into 9 and 10 loop islands. The separation between the islands is given by the length of the bars in the 0-homology graph. Note that these sets of points are the first cases that we find with $\mathcal{E} > 0$. Although the first set (salmon) has a larger $\mathcal{N}$, the parameter $\mathcal{E}$ is larger for the second set (brown), indicating that the loop islands are more persistent for the last set, i.e., the brown loop islands are

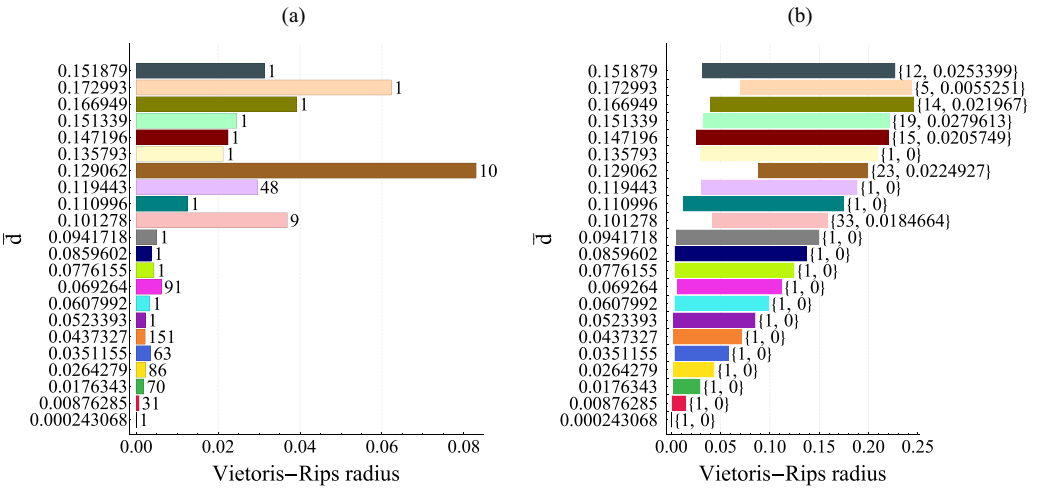

FIG. 6. TDA for Ra $= 10^4$. The bars correspond to each of the 22 Poincaré sets of points of the Lagrangian orbits studied. See Fig. 4(b). The color of the bars matches the color of the sets of points. The average distance $\bar{d}$ is shown in the column on the left-hand side. The left panel shows the longest bars (VR_0) of the 0-homology analysis, and the number of components in each set of points (p) is indicated at the end of each bar. The right panel shows the last bars for the 1-homology analysis and the parameters $\{\mathcal{N}, \mathcal{E}\}$ are given on the right-hand side of each bar.

"thicker" than salmon loop islands. The nonmonotonic $\bar{d}$ and the Vietoris-Rips radii where the bars in the 1-homology graph die show that the last four sets of points are mixed, see Fig. 4(b). This is consistent with the fact that $\mathcal{E} > 0$ for these cases.

The graph in Fig. 7(a) shows that the sets of points are arranged in three different groups. Group one includes the three sets with the smallest $\bar{d}$ and $\mathcal{E} = 0$. These sets are nested thin loops around the center c_3 (black dot). The parameters $p = 2$, $\mathcal{N} > 1$, and the large VR_0 indicate that the next three structures are clearly divided into two separate loop components; these constitute group two. The properties of these structures are similar to those described in Fig. 2(b).[1] The subsequent two bars with short VR_0's, $p = 1$ and $\mathcal{N} = 1$, belong to group one because the two sets organize themselves in thin loops with $\mathcal{E} = 0$ that enclose the structures with smaller $\bar{d}$. Their Lagrangian orbits form nested tori. Group three is characterized by large $\mathcal{E}$, $p = 1$, and the fact that $\bar{d}$ is not monotonously increasing. These properties indicate that the sets of points are scattered and do not form clusters. Therefore, the corresponding Lagrangian orbits are intertwined and exist in the exterior volume of the nested tori. This picture is consistent with Figs. 3(c) and 4(c). Case $\bar{d} = 0.198839$ features a very narrow bar where birth and death of the final 1-hole occur almost at the same Vietoris-Rips radii. This happens because the distribution of points in this set is such that the final hole is small and is not formed around the center c_3. This is clearly shown in video 7, $d = 0.198839$.avi, in Supplemental Material [29].

The set of points characterized by $\bar{d} = 0.084672$ (green) deserves a separate discussion. The TDA gives $p = 9$ and $\mathcal{E} = 0.00839467$, indicating that this set is formed by nine loop islands. A cursory inspection of Fig. 4(c) would lead to the conclusion that the islands are linear, but a more careful analysis reveals that the clusters are severely flattened loops. This is illustrated in Fig. 8, where the points set is redrawn with a zoom to appreciate that the clusters are indeed loop islands.

[1]The early birth and death of the $\bar{d} = 0.0560259$ in the 1-homology graph reflects the fact that the two kidneylike structures, in this case, do not form a central 1-hole as shown in video 7, $d = 0.0560259$.avi, in Supplemental Material [29].

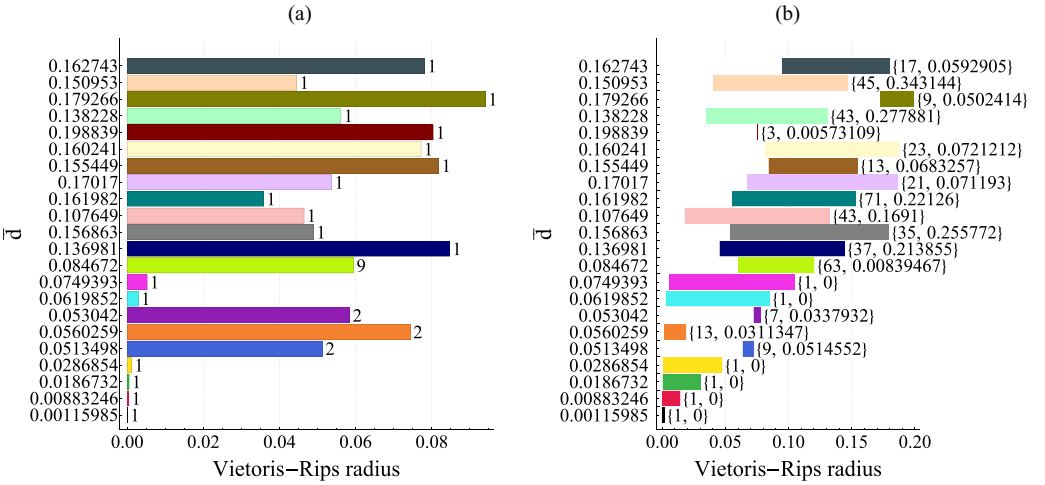

FIG. 7. TDA for $\mathrm{Ra} = 2 \times 10^5$. The bars correspond to each of the 22 Poincaré sets of points of the Lagrangian orbits studied. See Fig. 4(c). The color of the bars matches the color of the sets of points. The average distance $\bar{d}$ is shown in the column on the left-hand side. The left panel shows the longest bars (VR_0) of the 0-homology analysis, and the number of components in each set of points (p) is indicated at the end of each bar. The right panel shows the last bars for the 1-homology analysis, and the parameters $\{\mathcal{N}, \mathcal{E}\}$ are given on the right-hand side of each bar.

Here it is pertinent to remark that for the three cases previously described, the final bars in the 1-persistent homology correspond to the holes around the centers. For each set, the length of this bar gives a notion of how significant is the loop around the center c_i to characterize the distribution of dots.

In Figure 9, $\mathrm{Ra} = 5 \times 10^5$ the graphs and the parameters p, $\mathcal{N}$, and $\mathcal{E}$ show that there are two groups of sets of points. The first two sets belonging to the first group are nested thin loops around the center c_4 with small VR_0, $p = 1$, $\mathcal{N} = 1$, and $\mathcal{E} = 0$. The sets of points in the second group are characterized by large VR_0, $\mathcal{E} > 0$, and nonmonotonously increasing $\bar{d}$. These sets have nonuniformly distributed points, and given that $p = 1$, we conclude that no clusters are formed. The case $\bar{d} = 0.105527$ (yellow dots) comprises three ill-formed loop islands and points distributed in noneasily recognizable structures. This is shown on the left side of Fig. 10(a). The

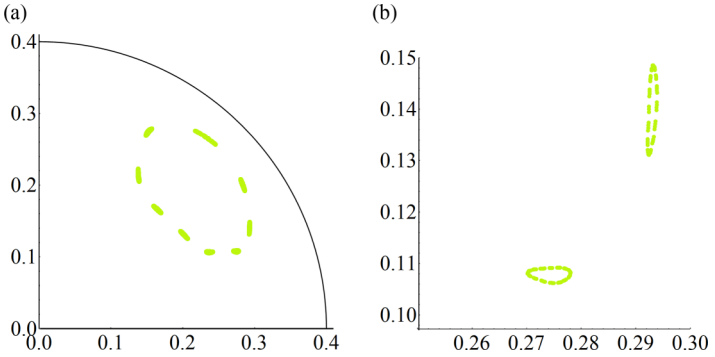

FIG. 8. (a) Poincaré map of the set of points with $\bar{d} = 0.084672$ for the case $\mathrm{Ra} = 2 \times 10^5$ (green points), redrawn from Fig. 4(c) to single out this case. (b) Amplification of the lower-right region of the set. The clusters are clearly loop islands.

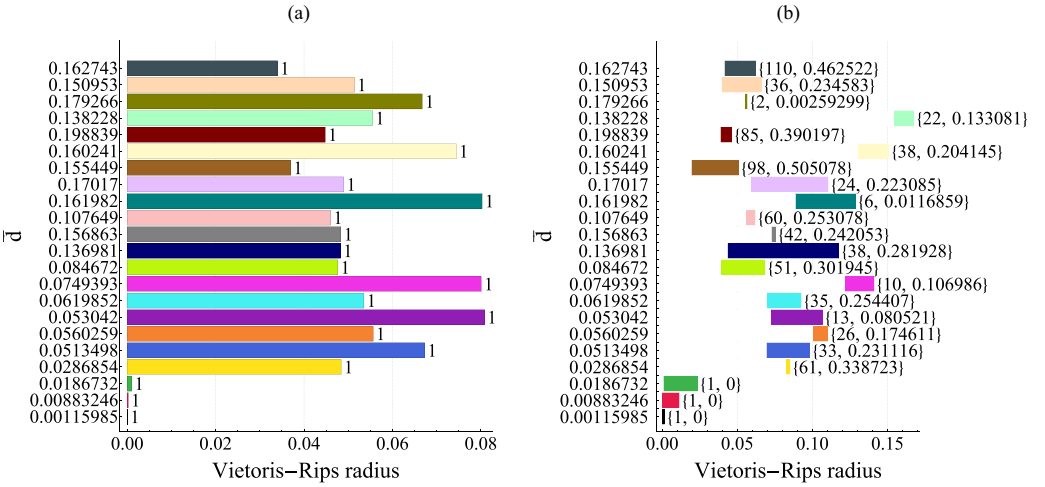

FIG. 9. TDA for $\mathrm{Ra} = 5 \times 10^5$. The bars correspond to each of the 22 Poincaré sets of points of the Lagrangian orbits studied. See Fig. 4(d). The color of the bars matches the color of the sets of points. The average distance $\bar{d}$ is shown in the column on the left-hand side. The left panel shows the longest bars (VR_0) of the 0-homology analysis, and the number of components in each set of points (p) is indicated at the end of each bar. The right panel shows the last bars for the 1-homology analysis and the parameters $\{\mathcal{N}, \mathcal{E}\}$ are given on the right-hand side of each bar.

case $\bar{d} = 0.155576$ (brown dots) has the largest $\mathcal{E}$ detected, indicating that many persistent holes are formed and filled as the Vietoris-Rips radii grow. Also notice that the 1-homology bar is born and dies with the smallest Vietoris-Rips radii in the second group. This property implies that in this set the points are distributed and do not form clusters. See Fig. 10(b).

In contrast to the previous three Rayleigh numbers analyzed, for $\mathrm{Ra} = 5 \times 10^5$, the last bar in the 1-persistent homology barcode does not correspond to the hole that surrounds the center c_4. An example of this phenomenon is the yellow dots set with $\bar{d} = 0.105527$ illustrated in video 9, $d = 0.105527$.avi, in Supplemental Material [29].

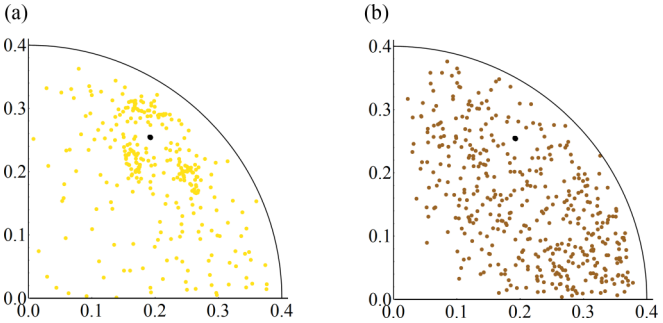

FIG. 10. Sets of points for $\mathrm{Ra} = 5 \times 10^5$; the center c_4 is represented by a black dot in each diagram. (a) $\bar{d} = 0.105527$ (yellow dots) and (b) $\bar{d} = 0.155576$ (brown dots).

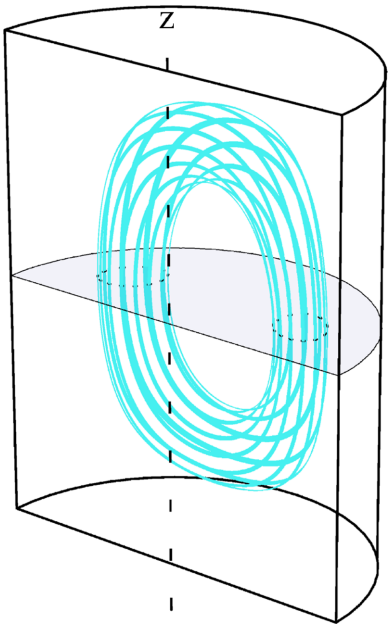

FIG. 11. Orbits that form the ribbon corresponding to points set with $\bar{d} = 0.0593613$ in Ra $= 8 \times 10^3$. The Poincaré map in the midplane ($z = 0.5$) is shown with black segments. This figure complements part of the information shown in Figs. 4(a) and 5(a).

C. Geometric properties of Lagrangian orbits

In this section, we take advantage of the information obtained in Section IV B to analyze the properties of the flow further. We incorporate in the analysis the sequential order of the Lagrangian orbits intercepts with the midplane, i.e., the order in which the points appear in the Poincaré maps.

Inspection of Figs. 3 and 4 shows that the trajectories that correspond to central points (black dots in Fig. 4) are, up to our precision, closed orbits. The sets of points in the Poincaré map that form loops around the center originated from Lagrangian orbits are embedded in toroidal surfaces (dense irrational frequency[2] orbits in KAM tori). Then, nested loops around central points are the intercepts of nested tori, see Fig. 3. These structures are related to the nested tori described by Arnol'd-Liouville Theorem for integrable area preserving Hamiltonians [11,38]. Also observed are orbits corresponding to sets of points that organize themselves in segments (linear islands). The flow of these orbits is arranged in a regular neighborhood of a torus knot which is contained in a toroidal surface, such a set resembles a ribbon as illustrated in Fig. 11.

Subsequent visits of the Lagrangian orbit to linear islands determine a turn of the ribbon in the toroidal direction; then, as explained in Sec. IV B, this is quantified by the parameter p. For each ribbon, the number of turns in the poloidal direction (q) coincides with the number of linear islands plus one between two segments visited by the Lagrangian orbit after one turn in the toroidal direction. We conclude that each ribbon determines an ordered pair (p, q) of two relatively prime numbers (if p, q are not relative primes, then not all linear islands are visited). These ordered pairs parametrize the resonances in the break-up of KAM tori. Notice that within the framework of the present analysis, sets divided into linear islands appear only in the Poincaré maps of the first two Rayleigh numbers studied. Table I contains the pairs (p, q) obtained in these cases.

[2]The formal definition of the frequency of an orbit can be found in the literature [11].

TABLE I. Pairs (p, q) of the sets divided in linear islands for $\mathrm{Ra} = 8 \times 10^3$ and 8×10^4. The color code corresponds to the sets shown in Figs. 4(a) and 4(b). Note that (p, q) pairs appear in ascending order with respect to $\bar{d}$.

Ra	(p, q)
8×10^3	(12, 1), (61, 5), (38, 3), (13, 1), (93, 7), (41, 3), (50, 3)
10^4	(31, 4), (70, 9), (86, 11), (63, 8), (151, 19), (91, 11), (48, 5)

Given that the pairs in Table I are relative primes, we find that the ribbons have similar geometrical properties to torus knots [39]. As noted above, ribbons can be identified with regular neighborhoods on the surface of the torus around the corresponding torus knots. Then, the sets of points in the Poincaré map that are divided into linear islands are determined by Lagrangian orbits that flow around torus knots.

Note that for $\mathrm{Ra} = 8 \times 10^3$, the ratios p/q in Table I are 12.00, 12.20, 12.67, 13.00, 13.28, 13.67, and 16.67, and for $\mathrm{Ra} = 10^4$, the ratios p/q are 7.75, 7.78, 7.82, 7.88, 7.95, 8.27, and 9.60. These sequences can be approximated by the expression

$$p/q = \lambda_0 + \lambda_1 \bar{d} + \lambda_2 \bar{d}^2. \tag{12}$$

The numerical value of the coefficients is given in Table II, and the corresponding plots are shown in Fig. 12.

As previously observed, for $\mathrm{Ra} = 10^4$ and $\mathrm{Ra} = 2 \times 10^5$ some sets of points with $\mathcal{E} > 0$ organize themselves in separated loop clusters (loop islands). On using similar criteria as those described in the first paragraphs of this section, we can identify the Lagrangian orbits that make up the loop islands as satellite knots over torus knots [39]. In the context of Hamiltonian dynamics, ribbons are known as resonances that wind themselves around KAM tori [11,40]. Our calculations show that the satellite knots found are trivial.

In Sec. IV B, we commented that the points set with $\bar{d} = 0.084672$ (green) in the case $\mathrm{Ra} = 2 \times 10^5$ is atypical (see Figs. 7 and 8). According to TDA, $p = 9$ and $\mathcal{E} = 0.00839467$. Given the relatively small $\mathcal{E}$, the Lagrangian orbit generating the set in the Poincaré map can be interpreted as a ribbon. The corresponding analysis yields $q = 4$ and $p/q = 2.25$, which is consistent with the descending trend of this ratio with increasing Rayleigh numbers. The fact that $\mathcal{E} > 0$ opens up the interpretation that this set is a satellite knot around a torus knot.

V. DISCUSSION AND CONCLUSIONS

A numerical study of the Lagrangian orbits in natural convection flows, together with the topological data analysis of the Poincaré maps generated by the orbits, is presented. The flows analyzed are assumed to take place inside a cylinder with an aspect ratio (height/diameter) of 1.25 and filled with a fluid with a Prandtl number of 6.67. We look into four cases of Rayleigh numbers, $\mathrm{Ra} = 8 \times 10^3, 8 \times 10^4, 2 \times 10^5, 5 \times 10^5$, where the flow is in a steady state. The TDA focused on the calculation of 0- and 1-persistent homologies. In Sec. IV B, we defined six useful parameters to classify the sets of points in the Poincaré maps, including the average distance of points set from the center $\bar{d}$ and $\mathcal{E}$, the parameter that quantifies the possible formation of holes and their dispersion in a specific set.

TABLE II. Numerical values of the coefficients in Eq. (12).

Ra	λ_0	λ_1	λ_2
8×10^3	12.8564	−40.6555	452.836
10^4	7.78445	−3.23415	153.934

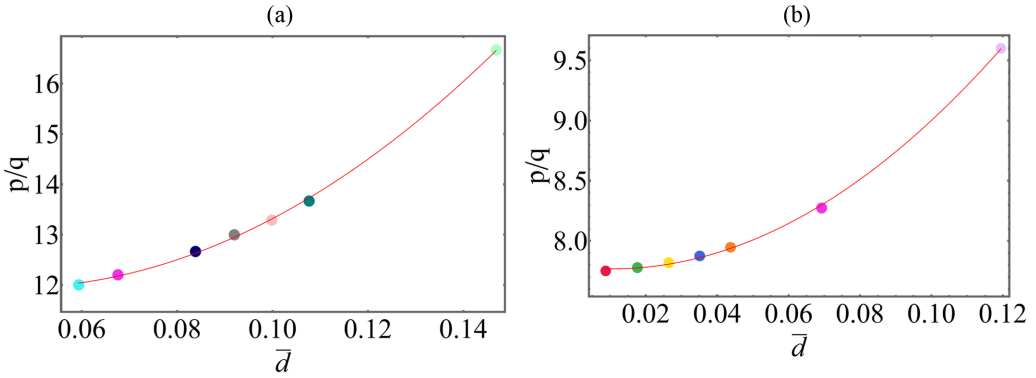

FIG. 12. The ratio p/q as a function of the average distance $\bar{d}$ for (a) $\mathrm{Ra} = 8 \times 10^3$ and (b) $\mathrm{Ra} = 10^4$. The color code of the dots corresponds to the sets of points in the Poincaré maps of Figs 4(a) and 4(b). The red lines are the fits given in Eq. (12).

Although the structures formed by the Lagrangian orbits have been described in the framework of Hamiltonian mechanics and Lagrangian advection, TDA gives accurate and quantitative information that facilitates the identification of features of the particular example analyzed that are predicted by the general theories. Also, this tool is used to find new information on the topological properties of sets of dots in the Poincaré sections.

For the smallest Rayleigh number ($\mathrm{Ra} = 8 \times 10^3$), the sets of points in the Poincaré map group themselves in continuous loops and in loops formed by separated p linear islands; these sets are defined by orbits embedded in nested tori. The linear island sets do not represent periodic orbits, and symplectic maps theory indicates that they have irrational frequencies [11]. The Lagrangian trajectory thus defines a dense set of points in the segments forming the loop. These sets were characterized by the pairs (p, q), which are identified with torus knots, and given the definition of q, they are related to the way in which the points are distributed in the loop. It is found that p/q is a quadratic function of $\bar{d}$ [see Eq. (12)], and then the existence of other (p, q) orbits for a specific $\bar{d}$ can be found by interpolation. More complete calculations (not shown in the figure) confirm this assertion.

As the Rayleigh number is increased ($\mathrm{Ra} = 10^4$), the linear islands also follow the trend described in the previous paragraph, and two different kinds of sets of points appear, loop islands (resonance zones) and irregularly distributed points (chaotic regions). For these sets, we find that $\mathcal{E}$ increases as the average distance $\bar{d}$ increases. Examples of orbits with periodic frequency can be located at the centers of loop islands. The information contained in Table I and condensed in Eq. (12) suggests that some sets of points in the Poincaré maps that visual inspection indicate are continuous lines are, in fact, formed by Lagrangian orbits with a relatively large number of clusters (p).

Interesting phenomena appear for $\mathrm{Ra} = 2 \times 10^5$. Near the center point, two kidneylike separated structures are formed. Taken separately, each cluster has a small $\mathcal{E}$ but, taken as two parts of a single system, as it should be because they belong to the same set of points, $\mathcal{E}$ is substantially larger. This indicates the formation of the persistent central hole. As explained in Sec. IV C, the set of points with $\bar{d} = 0.084672$ is divided into nine thin loop islands ($\mathcal{E} > 0$), which could be mistakenly interpreted as linear islands. This example opens up the possible interpretation that some of the sets classified as linear islands are thin loop islands. Most orbits found for the largest Rayleigh number analyzed, $\mathrm{Ra} = 5 \times 10^5$, have Poincaré sets of points with large $\mathcal{E}$ and are not embedded on identifiable two-dimensional manifolds with the present tools. It should be emphasized, however, that the parameters defined give properties of the chaotic sets; specifically, they indicate the absence of clusters and the formation and persistence of 1-holes.

The analysis contained in this study was possible because the numerical calculations provided highly precise data, and the topological data analysis reveals details that would be difficult or impossible to get with other methods. In particular, the quantification of geometrical properties of the sets of points in the Poincaré maps was accurately calculated with TDA.

APPENDIX: INTUITIVE INTRODUCTION TO TDA

TDA is a technique that uses simplicial complexes and tools in algebraic topology in order to extract features of the shape of a data cloud. Here we provide a brief and intuitive introduction to the necessary mathematical concepts and the general ideas used by TDA. For a more complete and formal discussion of the concepts, the reader is referred to monographs like Refs. [33,41]. An accessible introduction with less technical emphasis can be found in Refs. [42,43], and a reference with a specific focus on data science is Ref. [44]. Since we only analyze point clouds in the plane and for simplicity, we focus this introduction on the case of two-dimensional data clouds.

1. Approximating points clouds with Vietoris-Rips complexes

In order to analyze the shape of a point cloud with algebraic topology tools, first, it is necessary to approximate the data with some spaces where we can apply the corresponding tools. Here we introduce the Vietoris-Rips complexes, a widely used way to approximate data clouds.

We denote by $D_r(x) = \{y \in \mathbb{R}^2 : \|x - y\| \leqslant r\}$ the closed disk of radius $r > 0$ centered at the point x. Remember that the convex hull of a set $A \subset \mathbb{R}^2$ is the smallest convex set that contains A [41]. In particular, the convex hull of a finite set of points $\{x_1, x_2, \ldots, x_\ell\} \subset \mathbb{R}^2$ is a linear segment if the points are collinear or a filled polygon if they are not.

Suppose that $N = \{x_1, x_2, \ldots, x_\ell\} \subset \mathbb{R}^2$ is a point cloud. We denote by

$$r_{\min} = \min_{1 \leqslant i \neq j \leqslant \ell} \{\|x_i - x_j\|\} \quad \text{and} \quad r_{\max} = \max_{1 \leqslant i \neq j \leqslant \ell} \{\|x_i - x_j\|\},$$

the smallest and the largest distance between points of N, respectively. Note that if $r < \frac{r_{\min}}{2}$, then for every pair of points $x_i, x_j \in N$, the intersection $D_r(x_i) \cap D_r(x_j)$ is empty, and if $r \geqslant \frac{r_{\max}}{2}$, then for every pair of points $x_i, x_j \in N$, $D_r(x_i) \cap D_r(x_j) \neq \emptyset$.

Definition A.1 Let $r > 0$. The Vietoris-Rips complex of radius r of N, is the set $\mathcal{K}_r(N) \subset \mathbb{R}^2$ that is constructed by the union of the convex hulls of all the subsets $A \subset N$, such that, for all pairs $x_i, x_j \in A$, it is satisfied that $D_{\frac{r}{2}}(x_i) \cap D_{\frac{r}{2}}(x_j) \neq \emptyset$, i.e., $\|x_i - x_j\| \leqslant r$.

Notice that for $r < \frac{r_{\min}}{2}$, the Vietoris-Rips complex $\mathcal{K}_r(N)$ is equal to the set N, and for $r \geqslant \frac{r_{\max}}{2}$, $\mathcal{K}_r(N)$ is the convex hull of N. As an easy example, consider $N = \{(0, 0), (1, 0), (1, 1), (0, 1)\}$. Then, for $r_1 < \frac{1}{2}$, we have that $\mathcal{K}_{r_1}(N)$ is equal to N, for $\frac{1}{2} < r_2 < \frac{1}{\sqrt{2}}$, we have that $\mathcal{K}_{r_2}(N)$ are the four lines forming a square, and for $\frac{1}{\sqrt{2}} \leqslant r_3$, $\mathcal{K}_{r_3}(N)$ is the filled square with vertices at N.

Since there is no radius r such that the Vietoris-Rips complex $\mathcal{K}_r(N)$ is the one that best approximates the set N; then we construct a sequence of complexes as follows: Given two numbers $r_0 \leqslant \frac{r_{\min}}{2}$ and $s > 0$ such that $\frac{r_{\min}}{2} < r_0 + s$, we construct the sequence $r_j := r_0 + js$ with $1 \leqslant j \leqslant m$ and $r_{m-1} < \frac{r_{\max}}{2}$ and $r_m \geqslant \frac{r_{\max}}{2}$, then the Vietoris-Rips complexes corresponding to these radii form a nested sequence that satisfies

$$N = \mathcal{K}_{r_0}(N) \subset \mathcal{K}_{r_1}(N) \subset \mathcal{K}_{r_2}(N) \subset \cdots \subset \mathcal{K}_{r_{m-1}}(N) \subset \mathcal{K}_{r_m}(N),$$

where $\mathcal{K}_{r_m}(N)$ is the convex hull of N. This nested sequence of complexes is called a Vietoris-Rips filtration of N. An example is shown in Fig. 13; also, more examples of Vietoris-Rips filtrations are given in Supplemental Material [29].

2. Homology of spaces

Homology is a tool of algebraic topology that studies the shape of a space X. For every integer $i \geqslant 0$, there is a homology vector space $H_i(X)$ with coefficients in the field $\mathbb{Z}_2$ of integers

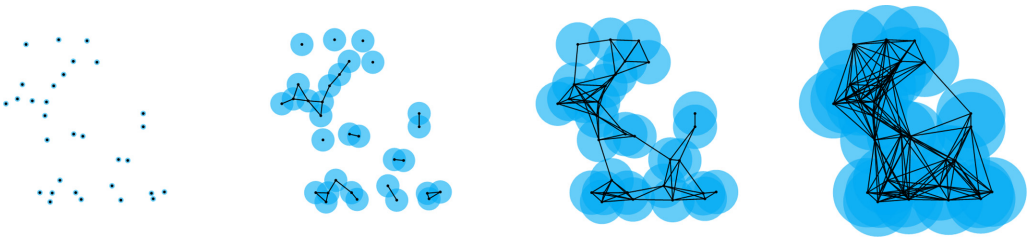

FIG. 13. The evolution of the Vietoris-Rips complexes as the radius r increases for a data cloud in $\mathbb{R}^2$. The first one is exactly the data cloud; the remaining are formed by joining convex hulls of some sets of points.

modulo 2. Roughly speaking, the elements in $H_i(X)$ are the i-holes in X and its dimension $\beta_i(X) := \dim[H_i(X)]$, is the number of "independent i-holes" in X. The number $\beta_i(X)$ is called the ith Betti number of X.

The formal mathematical definition of $H_i(X)$ is technical and it is not necessary for the purposes of this paper [41]. Here we only introduce an intuitive idea of the meaning of $H_0(X)$ and $H_1(X)$, which are the only relevant homologies for sets contained in the plane $\mathbb{R}^2$.

A 0-hole in X is one that can be "caught" with copies of the 0-sphere $\mathbb{S}^0 = \{x \in \mathbb{R} : \|x\| = 1\} = \{1, -1\}$ contained in X, i.e., unordered pairs p, q of points in X. A pair $p, q \in X$ that can be joined with a curve contained in X, does not catch a 0-hole since p can be moved toward q through such a curve (the same happens with q). Then, X has as many 0-holes as the number of "independent pairs of points in X that cannot be joined with a curve," i.e., $\beta_0(X)$ is the number of pieces into which X is divided (the pathwise connected components of X).

A 1-hole in X is a hole that can be "caught" with copies of the circle $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ contained in X, i.e., closed curves contained in X. A filled closed curve α contained in X, does not catch a 1-hole since the curve α can be continuously deformed to a point through the fill inside α. Then X has as many 1-holes as the number of "independent unfilled closed curves contained in X," the 1-Betti number $\beta_1(X)$.

As an example, we compute the number of 0- and 1-holes in the Vietoris-Rips complexes $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{K}_4$ in Fig. 13 listed from left to right.

(1) $\beta_0(\mathcal{K}_1) = 31$, which is exactly the number of points. $\beta_1(\mathcal{K}_1) = 0$ because there are no closed curves, and therefore, there are no 1-holes.

(2) $\beta_0(\mathcal{K}_2) = 12$, in the 12 components, there are points, segments, and unions of triangles with segments. Again $\beta_1(\mathcal{K}_2) = 0$ since there are four closed curves that are triangles, but they are filled.

(3) $\beta_0(\mathcal{K}_3) = 1$ because there is only one piece. In this case, $\beta_1(\mathcal{K}_3) = 2$ because there are a quadrilateral and a hexagon that are not filled, i.e., there are two 1-holes.

(4) Clearly $\beta_0(\mathcal{K}_4) = 1$ and $\beta_1(\mathcal{K}_4) = 1$ for the quadrilateral that is not filled. Notice that this 1-hole in $\mathcal{K}_4$ is different from the two 1-holes in $\mathcal{K}_3$.

Observe that convex sets are connected and do not have i-holes for all $i > 0$. Particularly, for convex sets, the 0- and 1-Betti numbers are 1 and 0, respectively.

3. Persistent homology and barcodes

Suppose that $N = \{x_1, x_2, \ldots, x_\ell\} \subset \mathbb{R}^2$ is a data cloud, and

$$N = \mathcal{K}_{r_0}(N) \subset \mathcal{K}_{r_1}(N) \subset \cdots \subset \mathcal{K}_{r_m}(N)$$

is a Vietoris-Rips filtration of N. The general idea of persistent homology consists of calculating the homology of each one of the Vietoris-Rips complexes $\mathcal{K}_{r_j}(N)$ with $0 \leqslant j \leqslant m$, and storing the information on how many of these complexes a hole appears, that is, for how many consecutive radii each hole persists.

Suppose that $\mathcal{K}_{r_i}(N) \subset \mathcal{K}_{r_j}(N)$. We proceed to analyze the behavior of the 0- and 1-holes with the growth of the radius.

(i) Let $\tau \in H_0[\mathcal{K}_{r_i}(N)]$, that is, τ represents a separate component of $\mathcal{K}_{r_i}(N)$. There are two options:

(1) $\tau \in H_0[\mathcal{K}_{r_j}(N)]$ and therefore, τ is a separate component of $\mathcal{K}_{r_j}(N)$. We say that τ persists at $\mathcal{K}_{r_j}(N)$.

(2) $\tau \notin H_0[\mathcal{K}_{r_j}(N)]$. Then τ is joined with another component, i.e., the 0-hole represented by τ disappears.

(ii) A closed curve $\sigma \subset \mathcal{K}_{r_i}(N)$ is also a closed curve in $\mathcal{K}_{r_j}(N)$. Let $\sigma \in H_1[\mathcal{K}_{r_i}(N)]$, that is, σ is a closed curve unfilled in $\mathcal{K}_{r_i}(N)$. There are two options:

(1) $\sigma \in H_1[\mathcal{K}_{r_j}(N)]$ and therefore, the closed curve σ is unfilled in $\mathcal{K}_{r_j}(N)$. We say that σ persists at $\mathcal{K}_{r_j}(N)$.

(2) $\sigma \notin H_1[\mathcal{K}_{r_j}(N)]$. The closed curve σ is filled in $\mathcal{K}_{r_i}(N)$.

Observe that $\beta_0[\mathcal{K}_{r_0}(N)] = \ell$ that is the number of points in N, and $\beta_1[\mathcal{K}_{r_0}(N)] = 0$ because there are no 1-holes; $\mathcal{K}_{r_m}(N)$ is the convex hull of $N \subset \mathbb{R}^2$ and therefore $\beta_0[\mathcal{K}_{r_m}(N)] = 1$ and $\beta_1[\mathcal{K}_{r_m}(N)] = 0$. Then the first two homologies satisfy: (a) 0-holes are *born* with the first radius r_0, and *die* when two or more connected components merge as the radius grows. This statement applies to all components except for the last one. (b) 1-holes are born for a certain positive value of the radius and die for a larger value but smaller than r_m.

Information obtained from the persistent homology of filtration is displayed in a barcode. Each bar in the barcode represents an element (σ or τ) in the homology and corresponds to increasing values of the filtration radius. In the one-dimensional space of the filtration radius, a bar begins when the element is born and ends when it dies. The length of a bar is called the persistence of the element.

Notice that the bars that have a significant persistence, represent holes that appear in the Vietoris-Rips complexes for many consecutive radii. This means that these holes correspond to real features in the shape of the data cloud N, see the examples in Sec. III. The Supplemental Material [29] contains eight videos that show the construction of barcodes for different data clouds.

[1] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability*. (Dover, New York, 1961).

[2] E. L. Koschmieder, Bénard Cells and Taylor Vortices. *Cambridge Monographs on Mechanics* (Cambridge University Press, Cambridge, UK, 1993).

[3] A. V. Getling, Rayleigh-Bénard Convection: Structures and Dynamics. *Advanced Series in Nonlinear Dynamics* (World Scientific, Singapore, 1998).

[4] M. Lappa, *Thermal Convection, Patterns, Evolution, and Stability* (Wiley, New York, 2010).

[5] G. Müller, G. Neumann, and W. Weber, Natural convection in vertical bridgman configurations, J. Cryst. Growth **70**, 78 (1984).

[6] G. Neumann, Three-dimensional numerical simulation of buoyancy-driven convection in vertical cylinders heated from below, J. Fluid Mech. **214**, 559 (1990).

[7] J. Núñez, M. López-Caballero, E. Ramos, G. Hernández-Cruz, M. Vargas, and S. Cuevas, Verification and experimental validation of a numerical simulation of natural convection in a slender cylinder, Int. J. Heat Fluid Flow, **38**118 (2012).

[8] G. Ramírez, J. Núñez, G. Hernández-Cruz, and E. Ramos, Natural convective three-dimensional flow structure in a cylindrical container, Int. Commun. Heat Mass Transf. **116**, 104616 (2020).

[9] M. F. M. Speetjens, G. Metcalfe, and M. Rudman, Lagrangian transport and chaotic advection in three-dimensional laminar flows, Appl. Mech. Rev. **73**, 030801 (2021).

[10] H. Aref, J. R. Blake, M. Budisic, S. S. S. Cardoso, J. H. E. Cartwright, H. J. H. Clercx, K. El Omari, U. Feudel, R. Golestanian, E. Gouillart, G. F. van Heijst, T. S. Krasnopolskaya, Y. Le Guer, R. S. MacKay,

V. V. Meleshko, G. Metcalfe, I. Mezic, A. P. S. de Moura, O. Piro, M. F. M. Speetjens, R. Sturman, J. Thiffeault, and I. Tuval, Frontiers of chaotic advection, Rev. Mod. Phys. **89**, 025007 (2017).
[11] J. D. Meiss, Symplectic maps, variational principles, and transport, Rev. Mod. Phys. **64**, 795 (1992).
[12] J. D. Meiss, Thirty years of turnstiles and transport, Chaos **25**, 097602 (2015).
[13] P. S. Contreras, M. F. M. Speetjens, and H. J. H. Clercx, Lagrangian transport in a class of three-dimensional buoyancy-driven flows, J. Fluid Mech. **832**, 5 (2017).
[14] A. J. Lichtenberg and M. A. Lieberman, Regular and Stochastic Motion, *Applied Mathematical Sciences*, Vol. 38 (Springer, New York, 1983).
[15] E. A. Jackson, *Perspectives of Nonlinear Dynamics* (Cambridge University Press, New York, 1989).
[16] P. S. Contreras, I. Ataei-Dadavi, M. F. M. Speetjens, C. R. Kleijn, M. J. Tummers, and H. J. H. Clercx, Topological equivalence between two classes of three-dimensional steady cavity flows: A numerical-experimental analysis, Phys. Fluids **31**, 123601 (2019).
[17] A. R. Sanderson, G. Chen, X. Tricoche, D. Pugmire, S. Kruger, and J. Breslau, Analysis of recurrent patterns in toroidal magnetic fields, IEEE Trans. Vis. Comput. Graph. **16**, 1431 (2010).
[18] B. Ravu, M. Rudman, G. Metcalfe, D. R. Lester, and D. V. Khakhar, Creating analytically divergence-free velocity fields from grid-based data, J. Comput. Phys. **323**, 75 (2016).
[19] J. Tierny, *Topological Data Analysis for Scientific Visualization* (Springer, Cham, 2018).
[20] R. Peikert (Ed.), *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications* (Springer, Berlin, 2012).
[21] H. Morita, M. Inatsu, and H. Kokubu, Topological computation analysis of meteorological time-series data, SIAM J. Appl. Dynam. Syst. **18**, 1200 (2019).
[22] J. R. Tempelman and F. A. Khasawneh, A look into chaos detection through topological data analysis, Physica D **406**, 132446 (2020).
[23] K. Mittal and S. Gupta, Topological characterization and early detection of bifurcations and chaos in complex systems using persistent homology, Chaos **27**, 051102 (2017).
[24] F. T. Pokorny, M. Hawasly, and S. Ramamoorthy, Topological trajectory classification with filtrations of simplicial complexes and persistent homology, Int. J. Robot. Res. **35**, 204 (2016).
[25] H. K. Moffatt, G. M. Zaslavsky, P. Comte, and M. Tabor (Eds.), *Topological Aspects of the Dynamics of Fluids and Plasmas* (Springer Netherlands, Dordrecht, 1992).
[26] R. I. Ricca, D. C. Samuels, and C. F. Barenghi, Evolution of vortex knots, J. Fluid Mech., **391**, 29 (1999).
[27] V. I. Arnold and B. A. Khesin, Topological methods in hydrodynamics, Annu. Rev. Fluid Mech. **24**, 145 (1992).
[28] M. Kramár, R. Levanger, J. Tithof, B. Suri, M. Xu, M. Paul, M. F. Schatz, and K. Mischaikow, Analysis of Kolmogorov flow and Rayleigh-Bénard convection using persistent homology, Physica D **334**, 82 (2016).
[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevFluids.7.123501 for further illustration of the methodology and details of the interpretation of TDA.
[30] R. Peyret, *Spectral Methods for Incompressible Viscous Flow* (Springer, New York, 2011).
[31] I. Mercader, O. Batiste, and A. Alonso, An efficient spectral code for incompressible flows in cylindrical geometries, Comput. Fluids **39**, 215 (2010).
[32] S. H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, *Studies in Nonlinearity* (Perseus Books, Cambridge, MA, 2001).
[33] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, RI, 2010).
[34] A. Zomorodian and G. Carlsson, Computing persistent homology, Discr. Comput. Geom. **33**, 249 (2005).
[35] B. T. Fasy, J. Kim, F. Lecci, C. Maria, D. L. Millman, and V. Rouvreau, TDA: Statistical Tools for Topological Data Analysis. R package version 1.7.7 (2021).
[36] K. Bajer, Hamiltonian formulation of the equations of streamlines in three-dimensional steady flows, Chaos Solitons Fract. **4**, 895 (1994).
[37] J. J. Valenzuela and S. E. Jerome, *Classical Dynamics: A Contemporary Approach*, 6th ed. (Cambridge University Press, Cambridge, UK, 2006).
[38] V. I Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1989).

[39] P. R. Cromwell, *Knots and Links*, 1st ed. (Cambridge University Press, Cambridge, UK, 2004).
[40] D. D. Nolte, *Introduction to Modern Dynamics: Chaos, Networks, Space and Time* (Oxford University Press, Oxford, 2015).
[41] Ž. Virk, *Introduction to Persistent Homology* (Založba UL FRI, Ljubljana, 2022).
[42] J. Tierny, Topological data analysis for scientific visualization, *Mathematics and Visualization* (Springer International Publishing, Cham, 2017).
[43] U. Fugacci, S. Scaramuccia, F. Iuricich, and L. De Floriani. Persistent homology: A step-by-step introduction for newcomers. In *Smart Tools and Apps for Graphics—Eurographics Italian Chapter Conference*, edited by G. Pintore and F. Stanco (The Eurographics Association, Eindhoven, The Netherlands, 2016).
[44] F. Chazal and B. Michel, An introduction to topological data analysis: Fundamental and practical aspects for data scientists, Front. Artif. Intell. **4**, 667963 (2021).