2012

Using machine learning to predict catastrophes in dynamical systems

Jesse Berwald

Tomas Gedeon

John Sheppard

Recommended Citation

Berwald, J., Gedeon, T., & Sheppard, J. (2012). Using machine learning to predict catastrophes in dynamical systems. Journal of Computational and Applied Mathematics, 236(9), 2235-2245.

Journal of Computational and Applied Mathematics 236 (2012) 2235–2245

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

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

Using machine learning to predict catastrophes in dynamical systems

Jesse Berwald [a,*], Tomáš Gedeon [a], John Sheppard [b]

[a] Department of Mathematical Sciences, Montana State University, MT, USA

[b] Department of Computer Science, Montana State University, MT, USA

ARTICLE INFO

Article history:
Received 20 November 2010
Received in revised form 1 September 2011

MSC:
65P99
37B30
68T05
37N99

Keywords:
Conley index
Dynamical systems
Machine learning

ABSTRACT

Nonlinear dynamical systems, which include models of the Earth's climate, financial markets and complex ecosystems, often undergo abrupt transitions that lead to radically different behavior. The ability to predict such qualitative and potentially disruptive changes is an important problem with far-reaching implications. Even with robust mathematical models, predicting such *critical transitions* prior to their occurrence is extremely difficult. In this work, we propose a machine learning method to study the parameter space of a complex system, where the dynamics is coarsely characterized using topological invariants. We show that by using a nearest neighbor algorithm to sample the parameter space in a specific manner, we are able to predict with high accuracy the locations of critical transitions in parameter space.

1. Introduction

Critical transitions are rapid changes in the behavior of nonlinear systems that arise after small changes in the parameters of a system. Examples of real-world systems that exhibit sudden shifts in their behavior include the Earth's climate, changes in ocean currents, collapses of plant and animal populations, and sudden swings in financial markets [1]. For instance, data indicate that the Earth's climate has swung between a "snowball" Earth and a "tropical" Earth many times in its history. These swings occur relatively rapidly on a geologic scale [2]. Also on a global scale, an abrupt change in the strength and direction of the Gulf Stream as a result of climate change would prove catastrophic for the European climate [3]. Evidence suggests that such a change was partially responsible for the three-hundred year Little Ice Age in Europe beginning in the 17th century [4]. Ecology provides another good source of examples of catastrophic change; for instance, eutrophication of a lake occurs when nutrient-rich pollution reaches a critical threshold, at which point water clarity is suddenly and greatly reduced due to a bloom of algae, which in turn kills submerged flora [1,5].

These systems are often modeled using systems of differential equations whose parameters are not known precisely, but for which parameter bounds can be inferred from the available data [6]. Critical transitions in the real world correspond to bifurcations in the models, where a change in parameter values results in a qualitative change in the behavior of the system.

The usual approaches used to predict critical transitions include studying the autocorrelation of a time series produced by the system, considering the skewness of the distribution of states, or testing whether the model "flickers" across a transition when stochastic forcing is added [1,5,7]. All of these are based on linearization analysis in a neighborhood of a generic local bifurcation, most often a saddle–node bifurcation. However, there are bifurcations where the local information about the equilibria does not change, but the rearrangement of the connecting orbits between equilibria, or birth of a periodic or

* Correspondence to: Department of Mathematics, College of William & Mary, VA, USA. Tel.: +1 4066001709.
E-mail addresses: berwald@math.montana.edu, jberwald@gmail.com (J. Berwald), gedeon@math.montana.edu (T. Gedeon), john.sheppard@cs.montana.edu (J. Sheppard).

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.11.006

homoclinic orbit, can lead to a critical transition. A change in the global structure of an invariant set is very difficult to detect computationally, and this challenge is compounded by the need to extend this computation over multi-dimensional parameter space.

Recent advances in theoretical and computational homology and Conley index theory rigorously allow one to summarize the full dynamics down to a predefined spatial scale using a combinatorial structure known as the Conley–Morse graph (CMG). To detect critical phase transitions, we sample a subset of the discretized parameter space and compute the CMG for each elementary unit in this subset. We then use the k-nearest neighbor (k-NN) algorithm, with $k = 1$, to cluster the regions in parameter space where the CMG's are isomorphic. We resolve the edges between these clusters by focusing the 1-NN algorithm on a subset of parameters near the edges. Using the nonlinear Leslie population model [8] as an example, we show that it is possible to sample and compute the dynamics for less than 20% of the parameters while correctly predicting the dynamics of nearly 95% of the remaining parameters.

We want to point out that although our methods may seem to be computing traditional bifurcation diagrams, a closer look reveals significant differences. Bifurcation structure can be very complicated [9–11] even for low dimensional systems. The regions of parameters where a system exhibits chaotic dynamics can be densely interlaced with regions that exhibit stable periodic orbits. This shows that in general it is impossible to compute a complete bifurcation diagram by a finite computation. Furthermore, standard approaches to computing bifurcation diagrams do not guarantee that the computed diagram is complete, i.e. that there are no further bifurcation curves inside the regions away from the computed bifurcation curves. We approach this fundamental limitation by using the finite computation to compute information that is coarser than the bifurcation diagram. However, we can guarantee that the information we compute is complete on our predefined level of resolution. We argue that this approach uses more effectively our fixed resource, which is the finite computation. Our interest here is in the detection of critical transitions on a large scale; thus, such characterization is ideal for helping us achieve this goal [12].

Furthermore, our method has a built-in robustness to changes in unmodeled parameters. If we assume that there are additional hidden parameters, then the computed information is the same for all systems sufficiently close to the given system in these parameters. Perturbations to which this applies include addition of a small bounded noise. The question of whether this also covers the addition of Gaussian (unbounded) noise is harder and is a subject of our current investigation.

2. Conley–Morse graphs

In the past decade, efficient computational topological methods have been developed to study dynamical systems [13–17]. We provide here a summary of the key points; a detailed description is beyond the scope of this paper. Central to the work described here is Conley index theory which computes topological invariants associated to isolated invariant sets from the index pairs of the corresponding isolated neighborhoods [18]. The Conley index can be applied to both continuous-time flows and maps, but since in our example we will use a discrete map, in our overview we focus on maps.

2.1. Morse decompositions

In this subsection, we describe how the dynamics of a system give rise to a Morse decomposition of the invariant set. This structure, together with the Conley index of the isolated invariant sets that form the Morse decomposition, is encoded in a Conley–Morse graph (CMG).

Let $f : X \to X$ be a *map* on a locally compact metric space X. That is, f is a continuous function, and for any given initial condition $x_0 \in X$, there is a unique trajectory $\{x_n\}$ where $x_n = f^n(x_0)$, $n \in \mathbb{N}$ or $\mathbb{Z}$ and $f^n := f \circ \cdots \circ f$ iterated n times. For the remainder of this section, we assume that f is invertible so that $n \in \mathbb{Z}$. We follow the exposition in [19] of Morse decomposition and Conley index. For more details, the reader is referred to the original paper. We begin with a few definitions that are basic to the rest of the section.

Definition 2.1. A subset $S \subset X$ is *invariant* under the map f if for all $x \in S, f(x) \in S$.

Let N be a compact subset of X. The maximal invariant set of N is the set of points that never leave N in forward or backward time,

$$\mathrm{Inv}(N, f) = \{x \in N \mid f^n(x) \in N,\ \forall n \in \mathbb{Z}\}.$$

We abuse notation and refer to $\mathrm{Inv}(N, f)$ (or $\mathrm{Inv}(N)$ when f is fixed) as the invariant set of N, as opposed the maximal invariant set. While invariant sets are in general not computable, isolated invariant sets, which we define next, are computable.

Definition 2.2. Let $N \subset X$, then

i. N is an *isolating neighborhood* if

$$\mathrm{Inv}(N) = \mathrm{Inv}(N, f) \subset \mathrm{int}(N);$$

ii. whenever $S = \mathrm{Inv}(N, f)$ and N is an isolating neighborhood, we say that S is an *isolated invariant set*,

where $\mathrm{int}(\cdot)$ denotes the interior of a set.

With the definition of an isolated invariant set in hand, we proceed by defining the related notion of *attractors* and *repellers* [20]. First, consider the long-term behavior of points under the action of f. The *forward orbit* of a point x_0, $\gamma^+(x) = \{f^n(x_0) \mid n > 0\}$, is the set of all iterates of x in forward time. For invertible f the *backward orbit* $\gamma^-(x)$ is defined analogously with $n < 0$.

Definition 2.3. Let $x_0 \in X$.

i. The point y is in the *omega-limit set* of x_0, denoted $\omega(x_0)$, if there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ of iterates in the forward orbit of f such that $\lim_{k\to\infty} f^{n_k}(x_0) = y$. In other words,

$$\omega(x_0) = \omega(x_0, f) = \bigcap_{n\in\mathbb{N}} \overline{\{f^k(x_0) \mid k \geq n\}}$$

where the bar signifies set closure.

ii. When f is invertible, the *alpha limit set* of x_0 is

$$\alpha(x_0) = \alpha(x_0, f) = \omega(x_0, f^{-1}).$$

The following proposition is a direct result of the compactness of X.

Proposition 2.1 (*Hale and Koçak [20]*)**.** *If X is compact, then $\alpha(x)$ and $\omega(x)$ are nonempty, compact, and invariant for every $x \in X$.*

Using Definition 2.3, we define the long-term behavior of points in the invariant set S.

Definition 2.4. For a compact, invariant set S, the subset $A \subset S$ is an *attractor* in S iff for each open neighborhood $U \supset A$,

$$\omega(U \cap S) = A.$$

Given the attractor A, the *dual repeller* R in S is

$$R := \{x \in S \mid \omega(x) \cap A = \emptyset\}.$$

With these definitions, we can decompose S into

$$S = A \cup R \cup C(R, A), \tag{1}$$

where $C(R, A)$ is the set of connecting orbits in S. That is, $C(R, A)$ is composed of points $x \in S$ such that $\omega(x) \in A$ or $\alpha(x) \in R$. The set (A, R) is called an *attractor–repeller* pair. Note that A and R are themselves isolated invariant sets.

Example 2.1. Consider the simple one-parameter system,

$$\begin{aligned} \dot{x} &= \lambda + x^2 \\ \dot{y} &= -y. \end{aligned} \tag{2}$$

When $\lambda < 0$ there are two equilibria $(\pm c, 0)$, where $c = \sqrt{|\lambda|}$. A saddle–node bifurcation occurs at $\lambda = 0$, at which point the two equilibria merge into a single equilibrium. For $\lambda > 0$, the system has no equilibria.

Fix $\lambda < 0$. Both equilibria lie in the interval $I = [-c - 1, c + 1]$. Then for $\kappa = c + 1$, $N = [-\kappa, \kappa] \times [-\kappa, \kappa]$ is an isolating neighborhood for I. The isolated invariant set is the line segment $S = \mathrm{Inv}(N) = [-c, c]$. Decomposing S into its attractor–repeller pair, we have $S = A \cup R \cup C(R, A)$, where $A = (-c, 0)$, $R = (c, 0)$, and $C(R, A)$ is the line segment between $(-c, 0)$ and $(c, 0)$.

The attractor–repeller decomposition can be generalized to a decomposition of the invariant set into a collection of partially ordered invariant sets. Let I be an index set and define $\mathcal{I} = (I, >)$ be a partially-ordered set (a *poset*) with the partial order $>$. We now define the Morse decomposition.

Definition 2.5. Consider a finite collection of disjoint, invariant subsets of S,

$$\mathcal{M}(S) = \{M(p) \mid p \in \mathcal{I}\}, \tag{3}$$

where for $M(p), M(q) \subset S$ we have that $M(p) \cap M(q) = \emptyset$ whenever $p \neq q$. We say that $\mathcal{M}(S)$ is a *Morse decomposition* of S if there exists a poset $\mathcal{I}$ such that

$$\forall x \in S \setminus \bigcup_{p\in\mathcal{I}} M(p)$$

there exist $p, q \in \mathcal{I}$, with $p > q$, for which $\omega(x) \in M(q)$ and $\alpha(x) \in M(p)$.

Example 2.2. The attractor–repeller decomposition for (2) is also a Morse decomposition,

$$\mathcal{M}(S) = \{M(i) \mid i = 0, 1;\ 1 > 0\},$$

where $M(0) = (-c, 0)$ and $M(1) = (c, 0)$. This is the attractor–repeller pair decomposition of S with $M(1) = R$ and $M(0) = A$. This is often represented as a graph. The elements of $\mathcal{M}(S)$ form the nodes, and the partial order amongst the nodes $M(p)$ determine the directed edges of the graph.

2.2. Conley index

In the previous subsection, we defined a partial order on a collection of disjoint isolated invariant sets by saying that $M(p)$ is "higher" than $M(q)$ if there is a solution that converges to $M(p)$ as $n \to -\infty$ and converges to $M(q)$ as $n \to \infty$.

Computationally, one can compute isolating neighborhoods of the Morse sets and these admit the same partial order [12,19]. An important related quality of isolating neighborhoods is that they *continue*. Suppose that we have a continuous parameterization of a map,

$$f_\mu : X \to X, \quad \mu \in [\mu_0, \mu_1],$$

where $\mu_0 < 0 < \mu_1$, and that we also have an isolating neighborhood N for f_0. Then there exists $\delta > 0$ for which N serves as an isolating neighborhood for f_μ whenever $|\mu| < \delta$. Thus, N is robust to perturbations. It follows that attractor–repeller pairs and Morse decompositions share the continuation property [19].

The Conley index of an isolated invariant set S is computed from an *index pair* (N, L) which is a pair of compact sets with $L \subset N$. Typically, but not always, the set N is an isolating neighborhood S [13,19].

Definition 2.6. Let S be an isolated invariant set. The pair of compact sets (N, L), $L \subset N$, forms an *index pair* for S under the following conditions:

i. $S = \mathrm{Inv}(\overline{N \setminus L})$ and $S \subset \mathrm{int}(N \setminus L)$.
ii. L is positively invariant in N: if $x \in L$ and $f^k(x) \in N$, then $f^k(x) \in L$ for all $k = 0, 1, \ldots, n$.
iii. Trajectories exit through L: given $x \in N$, if there exists a time K such that $f^K(x) \notin N$, then there is a value $n_0 \in \{0, 1, \ldots, K-1\}$ such that $f^{n_0}(x) \in L$.

Theorem 2.1 (*Mischaikow and Mrozek [19]*)**.** *Any isolated invariant set admits an index pair.*

Example 2.3. Consider the system in Example 2.1 for a fixed $\lambda < 0$, and the associated attractor–repeller pair $(A, R) = ((-c, 0), (c, 0))$. An isolating neighborhood for R is $N_R = [\delta, c+1] \times [-\kappa, \kappa]$, where $0 < \delta < c$. Let

$$L_R = (\{\delta\} \times [-\kappa, \kappa]) \cup (\{c+1\} \times [-\kappa, \kappa]).$$

Then (N_R, L_R) is an index pair. For A, an isolating neighborhood is $N_A = [-c-1, -\delta] \times [-\kappa, \kappa]$. Since there is no exit set, the index pair is (N_A, L_A) with $L_A = \emptyset$.

Definition 2.7. Let (N, L) be an index pair. The map $f : X \to X$ induces a continuous map $f_{(N,L)}$ on the quotient space $f_{(N,L)} : N/L \to N/L$, defined by

$$f_{(N,L)}(x) = \begin{cases} f(x) & \text{if } f(x) \in N \\ [L] & \text{otherwise,} \end{cases}$$

where $[L]$ denotes the equivalence class of points in L for which the following equivalence relation holds: $x \sim y$ iff $x = y$ or $x, y \in L$. This map in turn induces a homomorphism $f_* : H_*(N, L) \to H_*(N, L)$ on homology. The *homology Conley index* of an isolated invariant set is given by the shift equivalence class [21,13,19] of the pairs $(H_*(N, L), f_*)$.

Example 2.4. The homology Conley index for the invariant set $S_R = R$ in Example 2.3 is an equivalence class that contains

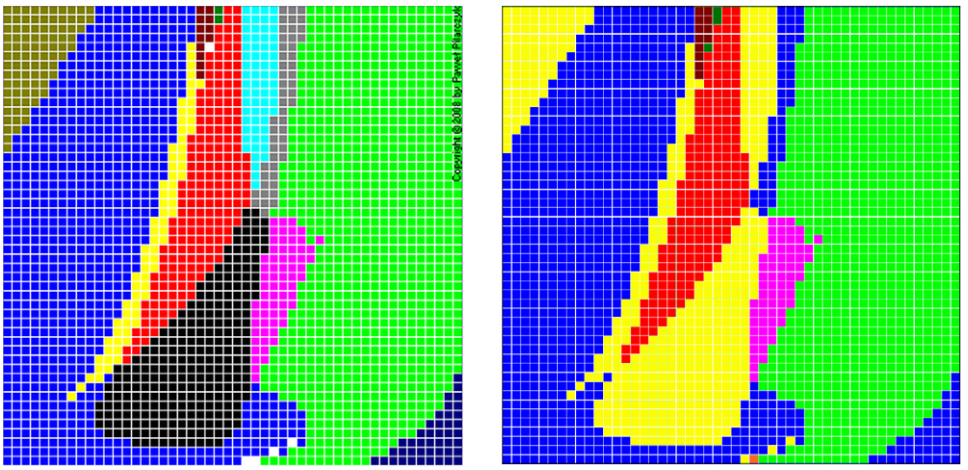
$$H_k(N_R, L_R) \equiv \begin{cases} \mathbb{Z} & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

with $f_* = id$ the identity map on the first homology group.

The Conley index information of the Morse sets along with the partial order induced by the Morse decomposition comprises a global description of a dynamical system (at a specific parameter) called the *Conley–Morse graph*. The oriented edges of the graph encode the partial order induced by Morse decomposition, while the vertices of the CMG correspond to Morse sets with the associated Conley index.

2.3. Advantage of using the Conley index

Observe that the invariant sets forming a Morse decomposition are isolated, and the corresponding isolating neighborhoods, index pairs, homology of the index pairs and the induced map on homology f_* are all computable. The only part of the Conley index that has not been algorithmically characterized is the shift equivalence class that forms the



(a) Original parameter space clusters. (b) Recomputed clustering.

Fig. 1. (a) The parameter space partitioned according to CMG graph classes determined using brute force methods. There are thirteen classes, including the uncolored grid elements. (b) Coarser partitioning of the parameter space computed using Morse graphs extracted from the data in [14]. Each color corresponds to identical Morse graphs and represents a distinct class for purposes of machine learning. There are eight classes in (b). Our algorithm compares its results to the classifications recorded in this image.
Source: Image taken from [14].

Conley index. Instead, one can compute the set of nonzero eigenvalues of the index map, which is an invariant of the shift equivalence (and thus the Conley index). This is not a complete invariant: if two matrices are shift equivalent, they have the same set of nonzero eigenvalues; the converse implication is not necessarily true. The key point is that the eigenvalue information is readily computable and can be used to conclude that equilibria, periodic orbits, even chaotic dynamics exist inside of a Morse set [22–24]. There are also results concerning the existence of connecting orbits and global bifurcations based on the behavior of the Conley index over sets of parameters [25,26].

The further advantage of using the CMG to characterize the global dynamics is the robustness of its construction in the parameter space, which is a consequence of the continuation property of the isolating neighborhoods. Choosing a spatial scale for both the parameter space and the phase space, the computed CMG is valid for each parameter value within the elementary unit of the discretized parameter space. Furthermore, if the CMG for overlapping grid elements of the parameter space are the same, and the corresponding isolating neighborhoods in the phase space pairwise intersect, one can conclude that the Morse decomposition over the union of the grid elements continues. This property has been used in [12] to construct continuation classes of CMG's over a parameter space.

Since we do not want to search the parameter space exhaustively, and thus will not sample overlapping grid elements of the parameter space, we will not use the pairwise intersection property to assign continuation classes. Rather, we will assign two grid elements to the same class if there exists a *directed graph isomorphism* between their corresponding CMG's. We take a moment to define the concept of a graph isomorphism. For a graph G, denote by $V(G)$ the vertices, and $E(G)$ the edges, of G. A directed edge is defined by $(u, v) \in E(G)$, where the order of the tuple implies an edge beginning at u and ending at v.

Definition 2.8. Let G and H be directed graphs. A *directed graph isomorphism* between G and H is a bijection f that maps $V(G)$ to $V(H)$ and $E(G)$ to $E(H)$ such that each $(u, v) \in E(G)$ is mapped to $(f(u), f(v)) \in E(H)$.

We have made the decision to ignore the Conley index eigenvalue information, since including this information would be computationally intensive and, in most important phase transitions the change in eigenvalue information is accompanied by a change in graph structure as well. Since we do not fully take into account Conley index information, our classes are coarser than the continuation classes and, consequently, we only detect the most significant phase transitions. As we explain in more detail in the next section, Fig. 1(a) and (b) compare the resulting classifications of parameter space using the Conley index eigenvalue information with the coarser clustering computed using only the graph.

A finite precision in both phase and parameter space constrain the conclusions that can be made about the underlying dynamics. When two CMG's computed on a fixed spatial scale are identical, we cannot conclude that their dynamics are homeomorphic, since bifurcations and changes in dynamics may happen on a spatial scale below the resolution of the calculation. However, it is our philosophy that such small perturbations are indistinguishable from noise.

2.4. Interpreting the Conley index

We finish Section 2 by describing how one can use Conley index information to interpret the type of the phase transition, even though we are not using this information to find the phase transitions. To illustrate this procedure, we examine Conley

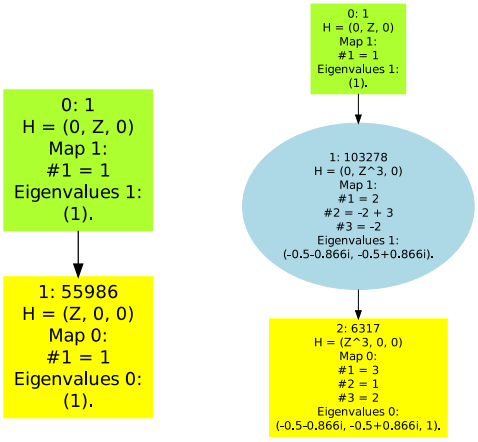


Fig. 2. (Color online) Two non-isomorphic CMG's are seen above. The CMG in (a) indicates that there is one repeller (top) and one attractor (bottom) in the system. The CMG in (b) indicates that the system has two repellers and one attractor (bottom). The eigenvalue information about the Morse sets, which we do not use in this paper, is encoded inside the nodes.
Source: Recomputed from data in [14].

index information presented in Fig. 2(a) and (b). These represent two parameter values in different regions in Fig. 1, which we explain in the next section.

On the graph in Fig. 2(a) the upper square represents the repeller whose homology H_1 is nontrivial (Z in the second slot of the vector). The induced map on the first homology is the identity with unique eigenvalue 1. Thus it has the same index as a fixed point with a one-dimensional unstable manifold [23]. The bottom square represents the attractor and the homology of the isolating neighborhood is nontrivial on level 0. This Morse decomposition is consistent with Example 2.3, but since we work on a fixed resolution in the phase space, we cannot be sure that both the attractor and the repeller consist of a single equilibrium.

In Fig. 2(b), the Morse decomposition has three sets. The top set contains the same information as the top set in Fig. 2(a). The middle set has three generators on first homology and the eigenvalues of the induced maps are $-\frac{1}{2} \pm \frac{\sqrt{3}}{2} i$. This represents a rotation by 120° and shows that the induced map on homology permutes the three generators in a circular fashion. The bottom set in the Morse decomposition has three permuting generators on the zeroth homology level. We interpret the phase transition between Fig. 2(a) and (b) as a bifurcation where a period three attractor and a period three repeller emerge from a saddle–node bifurcation. The eigenvalue 1 on the zeroth level of the bottom set shows that the attractor in Fig. 2(a) persists to Fig. 2(b), and the new pair of period three points is responsible for the complex pair of eigenvalues $-\frac{1}{2} \pm \frac{\sqrt{3}}{2} i$.

3. Problem studied

We illustrate our method on a nonlinear Leslie population model [8]. In addition, the partition of parameter space described, while relatively coarse, is well suited to situations in which the knowledge of parameter values is sparse, or where computation involving a family of high-dimensional parameters is expensive.

In the Leslie model, the population is partitioned into d generations, each with population $x_1, \ldots, x_d$, and a reproduction rate is associated to each generation. The nonlinearity comes from the assumption that the fertility decreases exponentially with the total size of the population. This situation is described by the map $g : \mathbb{R}^d \to \mathbb{R}^d$ which is defined by

$$\begin{pmatrix} x_1^n \\ x_2^n \\ \vdots \\ x_d^n \end{pmatrix} \mapsto \begin{pmatrix} (\theta_1 x_1^n + \cdots + \theta_d x_d^n) e^{\alpha(x_1^n + \cdots + x_d^n)} \\ p_1 x_1^n \\ \vdots \\ p_{d-1} x_{d-1}^n \end{pmatrix}, \tag{4}$$

where θ_j is the reproduction rate of generation j ($j = 1, 2, \ldots, d$) and p_i is the proportion of individuals from generation i that survive into generation $i + 1$ ($i = 1, 2, \ldots, d - 1$). Superscripts refer to the iteration of the map. Nonlinear models such as this have become increasingly important. For instance, the best linear models showed serious shortcomings after the 2000 US Census [8]. When using a more complicated nonlinear model, it becomes crucial to understand the bifurcations that arise in the $2d$-dimensional parameter space.

In this paper, we report the results of testing our method on a two-generation nonlinear Leslie population model over two-dimensional parameter space, which has been studied in [12]. As explained in the previous section, we sample the coarsened parameter space, compute CMG at these parameter values, and classify CMG's into clusters with identical CMG graphs. The boundaries between these clusters in the parameter space are predictions of phase transitions. Since [12] has performed an exhaustive search of a subset of the parameter space, we compare our classifier with the "ground truth"

provided by their work. We have reproduced their results in Fig. 1(a). Following [12], we define the map $g : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$g(x_1, x_2; \theta_1, \theta_2) = \begin{pmatrix} (\theta_1 x_1 + \theta_2 x_2) e^{-0.1(x_1+x_2)} \\ 0.7 x_1 \end{pmatrix}$$

where (θ_1, θ_2) are the reproduction rate parameters. The probability of intergenerational transfer and the exponential rate of fertility decay are fixed at $p = 0.7$ and $\alpha = -0.1$, respectively. This system exhibits a wide range of behaviors depending on parameters and therefore is an ideal candidate for our purposes [8].

Complicated structure of the parameter space is evident in both (a) and (b) of Fig. 1. As detailed in [12], we consider a region $[8, 37] \times [3, 50]$ of the (θ_1, θ_2)-parameter space that has been discretized into a 50×50 grid. For each box, a CMG, which is valid for all parameter values in that box, has been computed. Grid elements of a single color indicate that their CMG's are identical. As mentioned above, the case where computation is performed for every grid element and the Conley index eigenvalue information utilized is referred to as the brute-force method, the results of which are shown in Fig. 1(a). The case in which only the graph portion of the CMG is utilized to classify the dynamics is shown in Fig. 1(b). Grid elements of the same color between Fig. 1(a) and (b) indicate identical graph structure of the CMG at the particular parameter value where the element is located. Since we do not utilize the eigenvalue information, our recomputed clustering of the parameter space is necessarily coarser. We test the machine learning classification algorithm on this latter parameter space by comparing its predictions to this version of the brute-force classification. Accuracy of our algorithm is measured by the proportion of misclassified points in the test set.

4. Algorithmic methods

The k-nearest neighbor (k-NN) algorithm is a method of pattern recognition and is a fundamental tool used in cluster analysis. The k-NN algorithm is especially useful when the distribution of the data is unknown [27]. In the parlance of supervised machine learning, we define the following two terms.

Definition 4.1. A *training set* is a collection of vectors in a multidimensional space for which a "label" or "classification" is known for each vector.

Definition 4.2. A *test set* is a collection of vectors in the same space, each of whose classifications is treated as unknown.

Given an element u in the test set, the k-NN algorithm queries the training set for the k closest neighbors of u and assigns to u the class occurring most frequently amongst these neighbors. In the situation where $k = 1$, this implies that the space will be divided into Voronoi cells based on the training data, and the test data will be assigned the identity of the center of the Voronoi cell to which they belong. We determine the nearest neighbors using the Euclidean distance metric.

Following Arai et al. [12], we define a grid on $P \subset \mathbb{R}^d$.

Definition 4.3. A *grid* $\mathcal{Q}$ on P is a collection of nonempty, compact subsets of P such that

(i) $P = \overline{\bigcup_{Y \in \mathcal{Q}} Y}$
(ii) $Y = \overline{\text{int}(Y)}, \ \forall Y \in \mathcal{Q}$
(iii) $Y \cap \text{int}(Y') = \emptyset, \ \forall Y \neq Y' \in \mathcal{Q}$
(iv) For a compact set $K \in P$, the set $\{Y \in \mathcal{Q} \mid Y \cap K \neq \emptyset\}$ is finite.

For an equivalent definition, see [28]. The rectangular grid scale is given by the side length of each rectangle; see Section 3 and Fig. 1. Note that since P is compact the cardinality of $\mathcal{Q}$, denoted $\#\mathcal{Q}$, is finite. We consider the parameter space $P = [8, 37] \times [3, 50]$ described in Section 3.

Let $\mathcal{T}$ be a training set containing M grid elements sampled from $\mathcal{Q}$, where $M \in \mathbb{N}$ and $0 < M < \#\mathcal{Q}$. For each grid element $A \in \mathcal{Q}$ we define CMG(A) to be the function that assigns to A its CMG. From the machine learning perspective CMG(A) *classifies* A. For instance, Fig. 2 shows the distinct output of the CMG($\cdot$) function for two grid elements in $\mathcal{Q}$.

Define the test set as $\mathcal{R} = \mathcal{Q} \setminus \mathcal{T}$. Applying the 1-NN algorithm to $\mathcal{R}$ and $\mathcal{T}$, we obtain the classification set of $\mathcal{R}$ which we call $C_{\mathcal{R}}$. We define $C_{\mathcal{R}}$ as

$$C_{\mathcal{R}} := \{CMG(A) \mid A \in \mathcal{R}\}.$$

The parameter space $\mathcal{Q}$ is clustered into Voronoi cells determined by the locations of the vectors in $\mathcal{T}$ and their classes. We summarize this in the procedure KNN_Partition. On line 4, we call a 1-NN subroutine, KNN. Note that any function that performs nearest neighbor clustering can be used for KNN 4.

KNN_Partition($\mathcal{Q}, n$)

1 $\mathcal{T} \leftarrow \{x_1, \ldots, x_n\} \in \mathcal{Q}$ ▷ Training set
2 $C_{\mathcal{T}} \leftarrow$ CMG($\mathcal{T}$) ▷ Compute the CMG for each grid element in $\mathcal{T}$
3 $\mathcal{R} \leftarrow \mathcal{Q} \setminus \mathcal{T}$ ▷ Form test set
4 $C_{\mathcal{R}} \leftarrow$ KNN($\mathcal{T}, C_{\mathcal{T}}, \mathcal{R}$) ▷ Cluster elements of $\mathcal{R}$
5 $P \leftarrow C_{\mathcal{T}} \cup C_{\mathcal{R}}$ ▷ Partition of parameter space
6 **return** P

Fig. 3(a) shows an example of this classifications for a training set of 250 examples selected uniformly at random, which is 10% of the total number of grid elements.

4.1. Edges

In order to find a better classification, we perform a refinement of an initial classification. This involves determining the edges of the Voronoi cells and adding a portion of the grid elements along the edges the training set. A grid element in $\mathcal{Q}$ resides on an edge if it does not share the same class with at least one of its four neighbors on each side. (We ignore elements that fall on the boundary of $\mathcal{Q}$, since an edge will be detected by an interior neighbor.) We form the edge set from the subset of all grid elements in $\mathcal{R}$ for which at least one neighbor is found to be of a different class. The set of grid elements constituting the edges between Voronoi cells is likely large. Therefore, in practice, since we want to maintain as small of a training set as possible we sample only a portion (10%) of the set of grid elements along the edges. This relatively small number of additional training instances we denote by $\mathcal{E}$.

In the procedure Zoom, we summarize the edge enhancement algorithm. The edge detection step is handled by the subroutine Interfaces on line 2. The sampling along edges can be accomplished in different ways. For instance, one may wish to be careful to avoid choosing two grid elements opposite one another on either side of a boundary. Similarly one could enforce a sampling rule dictating that chosen elements be at least a certain distance apart. We are currently investigating these and other more refined methods of sampling. In this paper, we consider the boundary regions with no constraints on the samples. The principle goal at this stage of our investigation is to confirm that the 1-NN algorithm is a viable tool for the study of critical transitions in complex systems.

After the sample of edge elements, $\mathcal{E}$, has been identified, we form a new, more refined training set,

$$\mathcal{T}' = \mathcal{T} \cup \mathcal{E}.$$

In addition, the classifications CMG(A), for $A \in \mathcal{E}$, are appended to the set of classifications for $\mathcal{T}'$. We must also remove $\mathcal{E}$ from the initial test set to obtain the modified test set $\mathcal{R}' = \mathcal{R} \setminus \mathcal{E}$. Finally, the 1-NN algorithm is applied to the vectors in $\mathcal{R}'$ using the new training set $\mathcal{T}'$. This produces a new clustering of the parameter space. (See lines 3–7 in Zoom.) The procedure can be iterated until a desired level of error is reached, $\mathcal{T}'$ reaches a predetermined maximum size, or a predetermined number of edge enhancements has been performed. For simplicity, we assume the latter in the Zoom algorithm below.

The effect on accuracy is apparent between the classifications after one application of 1-NN algorithms in Fig. 3(a), and the clustering seen after ten applications of the edge enhancement procedure in Fig. 3(b). Notice that it is possible to detect additional classes when classifying $\mathcal{E}$. (Compare the results in Fig. 3 to the "baseline" in Fig. 1(b).) We show in Section 5 that the numerical accuracy of the classification more than doubles using edge enhancements.

Zoom(P, $\mathcal{Q}$, n, n_z)

1 **for** $i \leftarrow 1$ **to** n_z **do** ▷ Perform n_z edge enhancements
2 $\quad\mathcal{E} \leftarrow$ Interfaces(P) ▷ Find a subset of edge elements in P
3 $\quad\mathcal{T} \leftarrow \mathcal{T} \cup \mathcal{E}$ ▷ Append $\mathcal{E}$ to training set
4 $\quad\mathcal{R} \leftarrow \mathcal{R} \setminus \mathcal{E}$ ▷ Remove $\mathcal{E}$ from set of test elements
5 $\quad C_{\mathcal{T}} \leftarrow C_{\mathcal{T}} \cup$ CMG($\mathcal{E}$) ▷ Update classifications
6 $\quad C_{\mathcal{R}} \leftarrow$ KNN($\mathcal{T}$, $C_{\mathcal{T}}$, $\mathcal{R}$)
7 $\quad P \leftarrow C_{\mathcal{T}} \cup C_{\mathcal{R}}$ ▷ Update partition

8 Return P

5. Experimental results

We describe the results of two distinct phases in our experiments. The first serves to establish a baseline. It involves simple clustering with a 1-NN algorithm. The second improves upon the first by resolving edges discovered in the first experiment as summarized in Section 4.

5.1. Sampling without edge enhancement

As described above, for each trial we begin forming $\mathcal{T}$. Let

$$M \in \mathcal{S} = \{50, 100, 150, 200, 250\},$$

and let $\mathcal{T}$ be composed of M elements in $\mathcal{Q}$ chosen uniformly at random, which are then labeled. The remaining $2500 - M$ grid elements determine $\mathcal{R} = \mathcal{Q} \setminus \mathcal{T}$. The vectors in $\mathcal{R}$ are then classified using the 1-NN algorithm to obtain $C_{\mathcal{R}}$. The prediction error ϵ for $C_{\mathcal{R}}$ is defined to be the proportion of elements in $\mathcal{R}$ that are misclassified,

$$\epsilon(\mathcal{R}) = \frac{1}{\#\mathcal{R}} \sum_{A \in \mathcal{R}} \Delta[\mathrm{CMG(A)}, C_{\mathcal{R}}(A)],$$

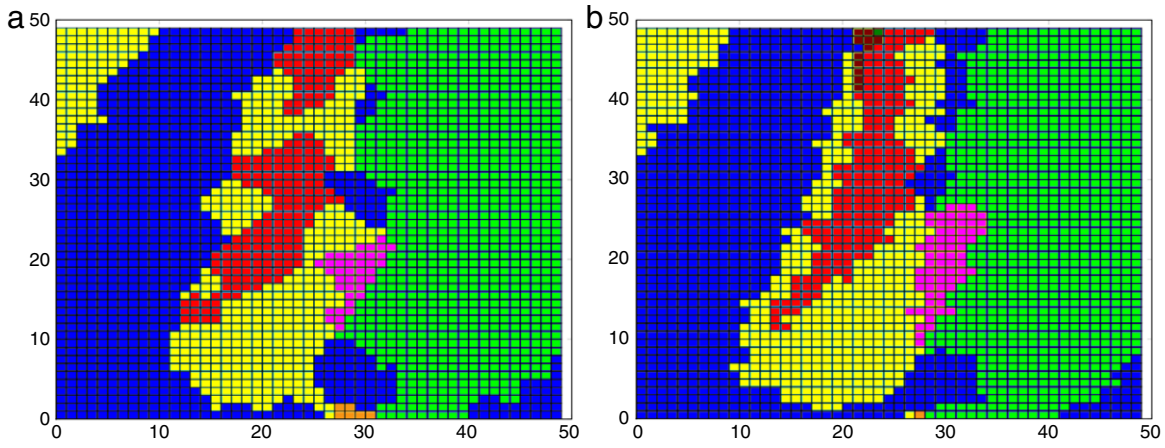


Fig. 3. (a) An initial 250 points are classified, with the remainder of classifications being assigned using the 1-NN algorithm. The mean classification error is slightly above 13.5% in this image. (b) The effect of ten edge enhancements, starting with the initial classification in (a). The mean classification error has been more than halved to approximately 6.8%. The total of all edge enhancements utilized only 241 additional examples. Colors correspond to the same classification of CMG's in Fig. 1(b).

where $C_{\mathcal{R}}(A)$ is the predicted class of A and CMG(A) is the true class of A. We use Δ to indicate when the predicted and true class do not match by defining it as

$$\Delta(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise.} \end{cases}$$

Recall that CMG(A) is determined from the "ground truth" classifications, shown in Fig. 1(b). Given $M \in \mathcal{S}$, we perform an ensemble of 1000 trials using training sets of size M. We then compute the mean prediction error over the ensemble of trials.

5.2. Zooming

The experiment described in 5.1 serves as a foundation for the edge resolution enhancement, or zooming, procedure. Following an initial clustering using the training set $\mathcal{T}$, we form an edge set as described in Section 4.1, which is a prediction of the location of the bifurcations (edges) in parameter space. As mentioned above, we now sample the grid elements defining these edges to form the set $\mathcal{E}$. The elements in $\mathcal{E}$ are classified and we obtain updated training and test sets, $\mathcal{T}'$ and $\mathcal{R}'$, respectively. We now apply the 1-NN algorithm to $\mathcal{R}'$. A more refined clustering of $\mathcal{Q}$ is obtained and the prediction error is computed.

5.3. Results

Based on the probabilistic nature of the problem, it is clear that by sampling more initial training points, the 1-NN classifier will better predict the bifurcations in the parameter space P. The numerical results presented below confirm this and describe the effects of edge enhancement on classification. For ensembles of 1000 trials beginning with training sets of initial size in $M \in \mathcal{S}$ for each trial, one would expect a decrease in the mean misclassification rate as the sample size increases. What is not clear, at first, is whether starting with training sets of different sizes and enhancing edges will continue to produce better predictions with lower misclassification rates. It could happen that there is a threshold M' such that given an initial training set of M' grid elements, the edges are already near optimal and little additional knowledge about the locations of bifurcations is gained by zooming in along the edges. We show that for the range of training set sizes used this is not the case. Fig. 4(a) shows the behavior of the mean classification error for the 1-NN algorithm over ten edge enhancements ("Zoom" in Fig. 4(a)) for each of the initial training set sizes. Indeed, there is still much room for improvement with 250 initial training examples. Note that after zooming ten times the mean classification error is still decreasing for all ensembles.

The comparison in Fig. 4(b) illustrates the most important aspect of zooming in on edges: after only five edge enhancements, the mean classification error for trials starting with 50 training examples reaches the accuracy of initially choosing 250 training examples. This accuracy is reached after training on a total of only 153 examples (50 initial training vectors plus 103 chosen over five edge enhancements). In Fig. 4(b), we emphasize the point at which an initial training set of 50 elements, plus zooming, reaches the initial accuracy of 250 training examples. Similar results hold for trials beginning with 100, 150, and 200 training examples. We conclude that the edge enhancement portion of the algorithm is able to compensate for a lower number of initial training examples and provide significant prediction improvement for all sample sizes.

We stopped the edge enhancement algorithm at ten iterations because of diminishing returns with respect to the number of training examples chosen. While further accuracy can be gained, the goal is to balance accuracy with minimal

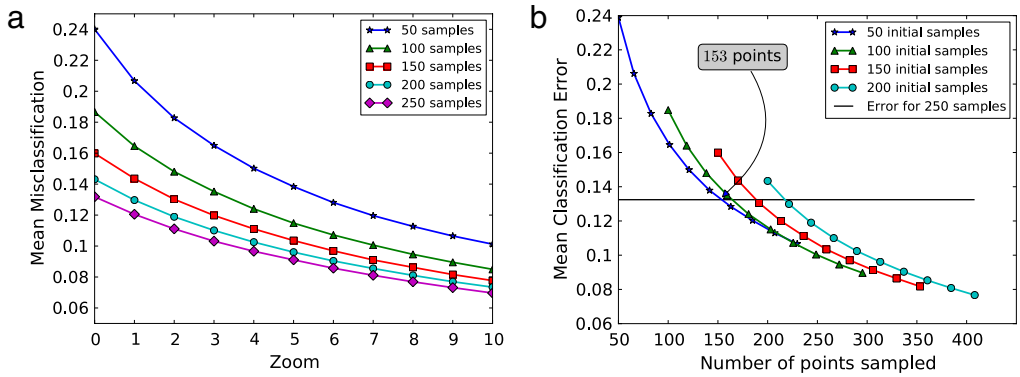


Fig. 4. (a) Mean classification error rates over 1000 trials for ten recursive edge enhancements. Initial sample sizes are in {50, 100, 150, 200, 250}. In all cases the variance was less than 0.0004. (b) Comparison of mean classification error for trials with initial samples in 50, 100, 150, 200 with respect to the mean error observed for training sets starting with 250 points. Markers denote the mean error for each zoom.

computational burden which we accomplish by choosing training samples judiciously. Returning to the specific example shown in Fig. 3, it is important to note that the increased accuracy seen from 3(a) to (b) results from ten edge enhancements that use less than twice the original number of training examples. The classification error for the initial prediction in Fig. 3(a) is 13.5% using 250 training examples. The prediction in Fig. 3(b) uses only a total of 491 training points—less than 1/5 of the available grid elements. The classification error is 6.8%, or rather 93.2% of points are classified correctly. Zooming in on probable bifurcations given by an initial sampling is a crucial step toward increasing accuracy while decreasing computational workload. Also of importance is that when starting with only six classes, we are able to identify the remaining two during the edge enhancement.

In this work, determining the edges (boundaries between the regions) is a relatively easy task since P is subset of $\mathbb{R}^2$. In a higher dimensional parameter space, this task could prove more difficult. However, for high dimensional parameter spaces, where a brute force approach to bifurcation analysis is prohibitive or impossible, detecting boundaries between the regions carries a comparatively small computational burden.

6. Conclusions

The prediction of bifurcations in dynamical systems that model complex real world systems is a critical area of study. For many real world systems including the climate, ecosystems and financial markets, the ability to predict a sudden change in behavior is crucial.

In this paper, we presented an approach which couples machine learning and Conley–Morse graphs to solve bifurcation prediction problems. A coarse, global characterization of the dynamics is valid over each elementary unit of a gridded parameter space and, by employing a straightforward nearest neighbor algorithm, we resolved the classification around potential bifurcations. In doing so, we showed that we can obtain between 90% and 94% accuracy of the global dynamics of a complex system using only a fraction of the grid elements in the discretized parameter space.

This work serves primarily as a "proof of concept" for the ideas described in the Introduction and Sections 3 and 4. We are currently in the process of extending our investigation along a number of paths, with the goal being an accurate method for bifurcation prediction in d-dimensional parameter spaces.

While effective in decreasing the misclassification, the edge enhancement used in the current implementation is not well suited to higher-dimensional parameter spaces. The continuation property of the Conley index can be used to aid in the choice of training elements. We are currently trying to use adaptive sampling techniques to improve our refinement around the interfaces. Simultaneously, we are working on extending our methods to the three-dimensional parameter space [14].

Many classification techniques exist in machine learning. One of the most straightforward to implement is the nearest neighbor algorithm. It is also well-suited to our specific partition problem. In conjunction with increasing the dimension of the parameter space, as well as the refinement of the sampling technique mentioned above, we are exploring the effects of k nearest neighbor clustering, along with other methods of classification that may be better-suited to higher-dimensional problems.

From a dynamical systems perspective, we also are pursuing the investigation of additional dynamical systems with complex parameter spaces. The main conceptual challenge is the application of the above techniques to real world problems where the state of the system is known only from a time series. Even a model may be missing, yet the classification of the behavior of an observed complex system is possible. While the first steps toward this goal fall within our framework, more needs to be done. Mischaikow et al. [29,30], use time series data to reconstruct the Conley index of an invariant set and its structure. The reconstruction of the entire Morse decomposition will likely require compilation of data from several time series samplings using different initial conditions and is subject to ongoing research.

Acknowledgments

JB and TG were supported in part by the DARPA FunBio program. JB would like to thank K. Michels for helpful discussions.

References

[1] Marten Scheffer, Jordi Bascompte, W.A. Brock, Victor Brovkin, S.R. Carpenter, Vasilis Dakos, Hermann Held, E.H. Van Nes, Max Rietkerk, George Sugihara, Early-warning signals for critical transitions, Nature (ISSN: 0028-0836) 461 (7260) (2009) 53–59. doi:10.1038/nature08227. URL: http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature08227.html.
[2] P.F. Hoffman, A neoproterozoic snowball earth, Science 281 (5381) (1998) 1342–1346. doi:10.1126/science.281.5381.1342. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.281.5381.1342.
[3] Michael Vellinga, Richard A. Wood, Global climatic impacts of a collapse of the Atlantic thermohaline circulation, Climate Change 54 (2002) 251–267. doi:10.1029/94JC00522. URL: http://www.sciencemag.org/cgi/content/abstract/278/5343/1582.
[4] L.D. Keigwin, The little ice age and medieval warm period in the Sargasso sea, Science (NY) (ISSN: 1095-9203) 274 (5292) (1996) 1503–1508.
[5] M. Scheffer, S. Carpenter, J.A. Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems, Nature (ISSN: 0028-0836) 413 (6856) (2001) 591–596. doi:10.1038/35098000.
[6] Michael Vellinga, Richard A. Wood, Jonathan M. Gregory, Processes governing the recovery of a perturbed thermohaline circulation in HadCM3, Journal of Climate (ISSN: 15200442) 15 (7) (2002) 764–780. doi:10.1175/1520-0442(2002)015(0764:PGTROA)2.0.CO;2. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0442(2002)015<0764:PGTROA> 2.0.CO;2.
[7] D.A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D.J. Frame, J.A. Kettleborough, S. Knight, A. Martin, J.M. Murphy, C. Piani, D. Sexton, L.A. Smith, R.A. Spicer, A.J. Thorpe, M.R. Allen, Uncertainty in predictions of the climate response to rising levels of greenhouse gases, Nature (ISSN: 1476-4687) 433 (7024) (2005) 403–406. doi:10.1038/nature03301.
[8] I. Ugarcovici, H. Weiss, Chaotic dynamics of a nonlinear density dependent population model, Nonlinearity 17 (2004) 1689–1711.
[9] M.S. Baptista, Celso Grebogi, E. Barreto, Topology of windows in the high-dimensional parameter space of chaotic maps, International Journal of Bifurcation and Chaos 13 (9) (2003) 2681–2688. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.2345&rep=rep1&type=pdf.
[10] Ernest Barreto, Brian Hunt, Celso Grebogi, James Yorke, From high dimensional chaos to stable periodic orbits: the structure of parameter space, Physical Review Letters (ISSN: 0031-9007) 78 (24) (1997) 4561–4564. doi:10.1103/PhysRevLett.78.4561. URL: http://link.aps.org/doi/10.1103/PhysRevLett.78.4561.
[11] J.A.C. Gallas, Structure of the parameter space of the Hénon map, Physical Review Letters 70 (18) (1993) 2714–2717. URL: http://link.aps.org/doi/10.1103/PhysRevLett.70.2714.
[12] Zin Arai, William Kalies, Hiroshi Kokubu, Konstantin Mischaikow, Hiroe Oka, Pawel Pilarczyk, A database schema for the analysis of global dynamics of multiparameter systems, SIAM Journal on Applied Dynamical Systems (ISSN: 15360040) 8 (3) (2009) 757. doi:10.1137/080734935. URL: http://link.aip.org/link/SJADAY/v8/i3/p757/s1&Agg=doi.
[13] Tomasz Kaczynski, Konstantin Mischaikow, Marian Mrozek, Computational Homology, Springer, 2004.
[14] Pawel Pilarczyk, Databases for the global dynamics of multiparameter systems, 2008. URL: http://chomp.rutgers.edu/database/.
[15] Vanessa Robins, Computational topology at multiple resolutions: foundations and applications, Ph.D. Thesis, University of Colorado, 2000. URL: http://lib.org.by/get/M_Mathematics/MP_Mathematicalphysics/RobinsV.ComputationaltopologyatmultipleresolutionsU.Coloradophdthesis,2000)(145s)_MP_.pdf.
[16] Daniel Wilczak, Computer assisted proofs in dynamics, 2010. URL: http://capd.ii.uj.edu.pl/.
[17] A. Zomorodian, G. Carlsson, Computing persistent homology, Discrete and Computational Geometry 33 (2) (2005) 247–274. URL: http://www.springerlink.com/index/6YCP1P7F0R6P1VRY.pdf.
[18] Konstantin Mischaikow, The Conley Index Theory: A Brief Introduction, Banach Center Publications, 1999, URL: http://www.springerlink.com/index/9k856x34kul46122.pdf.
[19] Konstantin Mischaikow, Marian Mrozek, Conley index theory, in: Handbook of Dynamical Systems II: Towards Applications, North-Holland, 2002, pp. 393–460 (Chapter 9).
[20] Anatole Katok, Boris Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge, 1995.
[21] John Franks, David Richeson, Shift equivalence and the Conley index, Transactions of the American Mathematical Society (2000) 1–23. URL: http://www.ams.org/tran/2000-352-07/S0002-9947-00-02488-0/S0002-9947-00-02488-0.pdf.
[22] Christopher McCord, Mappings and homological properties in the homology Conley index, Ergodic Theory and Dynamical Systems 8 (1988) 175–198.
[23] Christopher McCord, Konstantin Mischaikow, Marian Mrozek, Zeta functions, periodic trajectories, and the Conley index, Journal of Differential Equations 121 (2) (1995) URL: http://linkinghub.elsevier.com/retrieve/pii/S0022039685711291.
[24] Konstantin Mischaikow, Marian Mrozek, Chaos in the Lorenz equations: a computer-assisted proof, Bulletin of the American Mathematical Society (ISSN: 0025-5718) 32 (1995) 66–72.
[25] R.D. Franzosa, The connection matrix theory for Morse decompositions, Transactions of the American Mathematical Society 311 (2) (1989) 561–592. URL: http://www.ams.org/tran/1989-311-02/S0002-9947-1989-0978368-7/S0002-9947-1989-0978368-7.pdf.
[26] Christopher McCord, Konstantin Mischaikow, Connected simple systems, transition matrices and heteroclinic bifurcations, Transactions of the American Mathematical Society (ISSN: 00029947) 333 (1) (1992) 397. doi:10.2307/2154116. URL: http://www.jstor.org/stable/2154116?origin=crossref.
[27] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory (ISSN: 0018-9448) 13 (1) (1967) 21–27. doi:10.1109/TIT.1967.1053964. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1053964.
[28] Marian Mrozek, An algorithmic approach to the Conley index theory, Journal of Dynamics and Differential Equations (ISSN: 1040-7294) 11 (4) (1999) 711–734. URL: http://www.springerlink.com/index/G05J855470QW6606.pdf.
[29] K. Mischaikow, M. Mrozek, J. Reiss, A. Szymczak, Construction of symbolic dynamics from experimental time series, Physical Review Letters (ISSN: 0031-9007) 82 (6) (1999) 1144–1147. doi:10.1103/PhysRevLett.82.1144. URL: http://link.aps.org/doi/10.1103/PhysRevLett.82.1144.
[30] K. Mischaikow, M. Mrozek, A. Szymczak, J. Reiss, From time series to symbolic dynamics: an algebraic topological approach, Reconstructing Global Dynamics 15 (1997) 1–30.