Available online at www.sciencedirect.com
ScienceDirect
Physica D 223 (2006) 194–207
PHYSICA D
www.elsevier.com/locate/physd

Routes to chaos in high-dimensional dynamical systems: A qualitative numerical study

D.J. Albers[a,b,c,d,*], J.C. Sprott[b]

[a] Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany
[b] Department of Physics, University of Wisconsin, Madison, 1150 University Avenue, Madison, WI 53706-1390, United States
[c] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, United States
[d] Computational Science and Engineering Center, University of California, Davis, One Shields Ave, Davis, CA 95616, United States

Received 31 July 2004; received in revised form 28 April 2006; accepted 7 September 2006
Available online 9 October 2006
Communicated by C.K.R.T. Jones

Abstract

This paper examines the most probable route to chaos a high-dimensional dynamical systems function space (time-delay neural networks) endowed with a probability measure in a computational setting. The most probable route to chaos (relative to the measure we impose on the function space) as the dimension is increased is observed to be a sequence of Neimark–Sacker bifurcations into chaos. The analysis is composed of the study of an example dynamical system followed by a probabilistic study of the ensemble of dynamical systems from which the example was drawn. A scenario depicting the decoupling of the stable manifolds of the torus leading up to the onset of chaos in high-dimensional dissipative dynamical systems is also presented.

Keywords: Dynamical systems; Lyapunov exponents; Complex systems; Routes to chaos; Random matrix theory; Turbulence; Neural networks; Time-delay dynamics

1. Introduction

In their first edition of Fluid Mechanics [42], Landau and Lifschitz proposed a route to turbulence in fluid systems. Since then, much work in dynamical systems, experimental fluid dynamics and many other fields has been carried out concerning the routes to turbulence. In the early stages of such work, the terms turbulence and chaos were often used interchangeably. This usage has ceased because the subsequent development of the definitions of both chaos (which usually applies to a dynamical systems framework) and turbulence (which is a similar phenomenon, usually in a partial differential equations construction, yet with more implied structure [25]) have evolved and diverged. We will refrain from using the word turbulence since we will be studying discrete-time maps and not partial differential equations such as the Navier–Stokes equation. In this paper, we present results from a statistical study of the route to chaos in a class of high-dimensional, C^r dynamical systems on compacta. Our results contain both some reassurances based on a wealth of previous results and some surprises. We conclude that, for high-dimensional discrete-time maps, the most probable route to chaos (relative to the probability measure we impose on the class of dynamical systems) from a fixed point is via at least one Neimark–Sacker bifurcation, followed by persistent zero Lyapunov exponents, and finally a bifurcation into chaos. We observe both the Ruelle–Takens scheme as well as persistent n-tori, where $n \leq 3$ before the onset of chaos for state space dimensions up to and including 64.

The point of this report is two-fold. First, we propose a framework that consists of a manageable set of functions that can yield an understanding of high-dimensional routes to chaos in a practical sense. The goal of this portion is to

* Corresponding address: Max Planck Institute for Mathematics in the Sciences, Mathematical Physics, Inselstrasse 22, D-04103 Leipzig, Germany. Tel.: +49 341 9959 536.
E-mail addresses: albers@cse.ucdavis.edu (D.J. Albers), sprott@physics.wisc.edu (J.C. Sprott).
URLs: http://www.santafe.edu/~albers (D.J. Albers), http://sprott.physics.wisc.edu/sprott.htm (J.C. Sprott).

0167-2789/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.09.004

select a set of mappings that could form a bridge between real-world examples (the set of mappings we propose can approximate a very large set of real world examples) and rigorous mathematics. Second, we provide a survey of the routes to chaos in this set of mappings.

1.1. Background

To discuss the background necessary, we will start with the simplest standard construction for discussing routes to chaos in dynamical systems. Begin with an ordinary differential equation in R^k with a single real parameter μ, $\frac{dv}{dt} = F(\mu, v)$ where F is as smooth as we wish and $v \in U$ where $U \subset R^k$ is compact. At μ_0 there exists a fixed point, and at μ_c, $\mu_0 < \mu_c$, F is chaotic. The bifurcation sequence proposed by Landau [42] and Hopf [32] is the following: as μ is increased from μ_0 there will exist a bifurcation cascade of quasi-periodic solutions existing on higher and higher dimensional tori until the onset of "turbulence". In other words, the solutions would be of the following type, $x_{\mu_1}(t) = f(\omega_1, \omega_2)$, $x_{\mu_2}(t) = f(\omega_1, \omega_2, \omega_3), \ldots, x_{\mu_{k-1}}(t) = f(\omega_1, \omega_2, \ldots, \omega_k)$ for $\mu_i < \mu_{i+1}$, and where none of the frequencies are rationally related. However, Landau and Hopf's notion of turbulence was high-dimensional, quasi-periodic flow. Ruelle and Takens proposed both an alternative notion of turbulence, the strange attractor, and an alternative route to turbulence in their now famous paper [58]. Ruelle and Takens claimed that the Landau path was highly unlikely from a topological perspective. The basis for their claim originates in the work of Peixoto [52], who has shown that quasi-periodic motion on T^2 (the 2-torus) is non-generic[1] in the set of C^r vector fields. However, Peixoto's theorem applies only to flows on T^2 and not T^k for $k > 2$. There is a third perspective that uses a measure-theoretic notion of common instead of a topological notion. This perspective is born from a careful consideration of Poincaré sections of vector fields on a torus — rotations of the circle map. For diffeomorphisms of the circle, irrational rotations make up the full Lebesgue measure set of rotations, and suspensions of such diffeomorphisms correspond to quasi-periodic motion on T^2. Thus, it would seem that quasi-periodic motion of flows on T^2 would be high measure. However, there is not a one-to-one correspondence between flows on T^2 and discrete-time maps of the circle (e.g., the Reebs foliation or [51]). Moreover, quasi-periodic orbits of diffeomorphisms on the circle are structurally unstable and yet occupy the full measure of the dynamics. For flows on T^2, the structurally stable, hyperbolic periodic orbits are topologically generic; however, it is likely that quasi-periodic orbits are common in a measure-theoretic sense on T^2. Beyond an understanding of the most common dynamics is the more difficult issue of understanding the bifurcations between the various dynamical types upon slight parameter variation (i.e. quasi-periodic orbits bifurcating to strong-resonance periodic orbits). There remain many open questions regarding bifurcations of periodic orbits. The codimension 1 bifurcations (bifurcations involving one parameter or a single (pair of) real (complex) eigenvalue(s)) are well understood. However, there are many remaining problems regarding codimension 2 (and higher) bifurcations (the status of such problems can be found at the end of section (9.1), page 397 of [41]). How these many pieces will fit together in practice is unclear and comprises a good portion of motivation for our study.

Ruelle and Takens proved two results for flows relevant to this report. The first is a normal form theorem for the "second" Hopf bifurcation for vector fields, or the "first" Hopf bifurcation for maps (often referred to as the Neimark–Sacker bifurcation [59,46]). This theorem gives a normal form analysis of the bifurcation of an invariant circle of a flow, but it does not state the type of dynamic that will exist upon the loss of stability of the invariant circle. The second relevant result was that, given a quasi-periodic solution $f(\omega_1, \ldots, \omega_k)$ on T^k, $k \geq 4$, in every C^{k-1} small neighborhood of $f(\omega_1, \ldots, \omega_k)$, there exists an open set of vector fields with a strange attractor (prop. 9.2 [58]). These results were extended by Newhouse, Ruelle and Takens [48], who proved that a C^2 perturbation of a quasi-periodic flow on T^3 can produce strange (axiom A) attractors, thus reducing the dimension to three for tori with quasi-periodic solutions for which an open set of C^2 perturbations yield strange attractors. The basic scheme used by Takens, Ruelle and Newhouse was first to prove a normal form theorem from periodic orbits to 2-tori in vector fields and then to prove something about how the 2-tori behave under perturbations — showing that bifurcations of m-tori to $(m+1)$-tori will yield chaos since 3-tori can be perturbed away to axiom A chaotic attractors.

If the story were only as simple as a disagreement between topological and measure-theoretic viewpoints, and if codimension 1 bifurcations were the only bifurcations observed, it is likely that no problem would remain. However, out of the complexity of the dynamics and the difficulties posed by bifurcation theory regarding what happens to bifurcations of resonant periodic orbits and quasi-periodic orbits, the field of quasi-periodic bifurcation theory was born [13,10]. It is beyond the scope of this work to discuss the more recent and detailed history; for those interested, see [11,17,35,22,12,36,16], and [34]. The question regarding the most common route to chaos is, in any but a very select set of specific examples, still an open and poorly defined question. Even analytically piecing together the types of bifurcations that exist en route to chaos has been slow and difficult.

To achieve an understanding of routes to chaos in high-dimensional dynamical systems, while circumventing current abstract mathematical difficulties, several scientists have performed numerical experiments of a statistical nature. One of the early experiments was performed by Sompolinksky et al. [61], who analyzed neural networks constructed as ordinary differential equations. The primary goal of their construction was the creation of a mean field theory for their networks from which they would deduce various relevant properties. Their network architecture allowed them to make the so-called local chaos hypothesis of Amari, which assumes that the inputs are sufficiently independent of each other that

[1] A property is generic if it exists on subset $E \subset B$, where E contains a countable intersection of open sets that are dense in the original set B.

they behave like random variables. In the limit of infinite dimensions, they find regions with two types of dynamics, namely fixed points and chaos with an abrupt transition to chaos. In other words, the length of the parameter interval pertaining to the route to chaos decreases to zero as the size of the network is increased. We have found that, while the Euclidean length of the routes to chaos region decreases with increasing network dimension, the number of decades of the parameter pertaining to the routes to chaos region remains relatively constant. Doyon et al. [20,21] studied the route to chaos in discrete-time neural networks with sparse connections. They found that the most probable route to chaos was a quasi-periodic one regardless of the initial type of bifurcation. They justified their findings, which we will address in Section 4, with results from random matrix theory. Cessac et al. [15] came to similar conclusions with slightly different networks. They provided a mean-field analysis of their discrete-time neural networks in much the same manner as Sompolinsky et al. did for the continuous-time networks. It is important to note that these studies use a network architecture that is quite different from the architecture we will employ in this paper. We chose to use a time-delayed network; we were motivated by both the relative ease of computation for the derivative map and the approximations theorems of Hornik et al. [33]. What is reassuring is that many of the results for the different architectures are qualitatively similar despite the networks being fundamentally different.[2]

We contribute to the existing partial solution in the following ways: we will provide a framework for numerical analysis that does not have a priori tori built in — motivated by a marriage of the topological approximation results of Takens [65] and neural network approximation results of Hornik et al. [33]; and we will provide evidence that the most common route to chaos (relative to the probability measure we utilize), as the dimension is increased, consists of various bifurcations between periodic orbits of high-period and quasi-periodic orbits. Moreover, flip and fold bifurcations (due to real eigenvalues) occur with less and less frequency (relative to the measure we impose) as the dimension of the dynamical system is increased.

There have, of course, been many fluid experiments; but since our work is related more to the theoretical work of the aforementioned researchers, we will refrain from a summary and instead encourage the reader to consider, as a starting point, [22,63,29], and [64]. Establishing a relationship between experimental results in natural systems and a study such as the one presented here must be done using measure-theoretic language and will be briefly discussed in the section that follows.

1.2. The experiment

Ideally, we would study the C^r function space; however, in a computational format this is not possible because C^r does not admit a measure which is required for performing a numerical experiment. In lieu of this, we perform a statistical survey of a space of mappings (neural networks) that can approximate mappings and their derivatives from the C^r function space. The space of neural networks that we use here has been shown to be dense in C^r on compacta (as well as most Sobolev spaces) and admits a (probability) measure. However, for a function space to be truly dense in C^r on compacta, the mappings need infinitely many parameters — in particular the number of parameters must not have a dependence on the number of state-space variables. One of the motivations for choosing neural networks is the simplicity of the parameter space — the parameter space of the neural networks is the standard Euclidean space R^p where $p \in N$ is unconstrained by the number of state-space dimensions. Clearly an infinite parameter space is not achievable in numerical experiments; we will study spaces with the number of parameters ranging from 576 to 2113 (i.e. R^{576} to R^{2113}), which are not infinite, but not trivially small either. One of the key features is that each neural network can be identified with a point in R^p — the density of the space of neural networks in C^r is achieved as $p \to \infty$. This all yields a convenient framework in which to perform the numerical experiment — we impose a probability measure on the entirety of the space of neural networks via a measure on R^p. Thus, all mappings in the (finite-dimensional) space are represented measure-theoretically in the same way that all real numbers are represented by the normal distribution on R^1.

The experiment we perform is limited to an analysis of how the route to chaos varies as the dimension and number of parameters of the class of dynamical systems we utilize is increased. The analysis begins with a careful study of a particular example, followed by statistical results for an ensemble of dynamical systems. Due to the diversity of dynamics and bifurcation sequences yielded by the measure we impose on function space (again via the parameter space), we will analyze the example using bifurcation diagrams, phase plots, the Lyapunov spectrum, and the largest Lyapunov exponent. The example we will study is typical of the 500 cases (with the given number of parameters and dimensions) that we have observed in the sense that between the first bifurcation and the onset of chaos, the only type of orbit that exists is either quasi-periodic, or it is periodic with periods high enough such that they are indiscernible from quasi-periodic orbits. The statistical results that follow reinforce this viewpoint. It is nearly impossible to specify all the bifurcation sequences because the number of bifurcations between a fixed point and chaos increases with dimension (the number of bifurcation routes $> 3^{\text{number of bifurcations}}$). Instead, we will study the fraction of different bifurcations at each bifurcation. It is worth noting that there is a considerable difference between the observed routes to chaos at low and high dimensions.

This construction raises certain questions that are central to the proper interpretation of the results in this paper; of particular importance are questions regarding the relationship between results subject to the framework we employ and results subject to the frameworks of mathematics and science. It is not surprising that connections and relevance of the various frameworks hinge on statements in compatible languages — thus, it will be very useful to discuss the interaction of the

[2] For instance, the local chaos hypothesis is not valid for time-delayed networks, but many of the results of the aforementioned authors are similar.

various language sets. Starting with mathematics, the most obvious question is: how representative of C^r space is the finite-dimensional approximation we use? It is true that the entire (finite-dimensional) space of neural networks is included in our study via the probability measure on the parameter space — thus in an approximate sense we are studying a discretized representation of C^r space on compacta, but how can this be interpreted in a concrete way? To start with, there are two relatively absolutist means of comparing the results in this study with abstract mathematical results in a mathematical sense, measure-theoretically and topologically. In the construct used in this paper, both comparison methods are doomed; C^r does not yield a measure so a comparison in measure is impossible; the spaces we are studying are finite dimensional, and thus of infinite codimension (non-generic with respect to Baire category) with respect to C^r. This does not mean that it is impossible to compare results for C^r with results in this study; it just means that relationships with our results and the more abstract results of, say, structural stability theory will not be built with traditional mathematics language sets of commonality. Many of the fruitful mathematical dynamics results are with respect to certain assumptions, and the likelihood of these assumptions being met can be compared in measure with what is observed in the construction we are using. For instance, if, in our study, hyperbolicity is always observed, then the measure used is representative of systems that are likely structurally stable (note, hyperbolicity is not the only requirement for a dynamical system to be structurally stable, see [43,56,55]). Thus, the structure of the comparison statements will be in terms of probability of certain properties being met. From this, the applicability of the abstract mathematical results can be deduced. Analogous scientific questions addressing how representative this particular construction is of nature must also be carefully qualified, and several issues must be circumvented. While it is hoped that buried in C^r are representations of many natural phenomena (nearly every physical model is, in practice, an element of C^r), the direct relationship between the function space we are approximating to natural systems is yet unresolved. It goes without saying that this question is difficult to answer — in large part because of the lack of a measure-theoretic structure on both C^r and nature. The key to a comparison between the neural network construction and a specific natural system is again a more carefully worded, precise question: how relevant is the weight construction we employ to phenomenon X? This question, theoretically, is answerable. One could train an *ensemble* of the neural networks we will introduce in Section 3 on data from phenomenon X, and then compare the measures on the parameters captured on the training set with the measures we impose. From this a measure-theoretic analysis can be performed. For example, if the measures are singular with respect to each other, then there is no observable link between phenomenon X and our construction. At the heart of comparisons between studies such as the one in this paper and analysis of natural or abstract mathematics systems is the ability to use the language of measure-theoretic probability theory. In particular, the mathematical language for studying the space of measures on, say, R^p, is information geometry [6]. Before undertaking such a comparison we must first understand the effects of imposing a concrete measure on the space of neural networks, and that is precisely the aim of this current work. For readers interested in a more complete and mathematically precise description of this framework, see [2,50].

2. Lyapunov spectrum

A primary tool of analysis will be the Lyapunov characteristic exponents (LCE) because they are a good measure of the tangent space of the mapping along its orbit, and because they reveal the global geometry of the attractor. For a d-dimensional system, the spectrum consists of d LCEs: $\chi_1 \geq \chi_2 \geq \cdots \geq \chi_d$, where indexing is chosen to give a monotonic ordering. In a computational setting, a positive largest Lyapunov exponent (LLE) is the hallmark of chaos. Ruelle (as well as Katok and Pesin) proved that negative Lyapunov exponents correspond to global stable manifolds or contracting directions, and positive Lyapunov exponents correspond to global unstable manifolds or expanding directions [57,38,53]. Situations with neutral directions are significantly more difficult from a theoretical standpoint; yet, in general, a zero exponent corresponds to a neutral direction (see [14]). Computationally, when the largest Lyapunov exponent for a discrete-time map is zero and all other exponents are less than zero, there exists a neutral direction — a quasi-periodic orbit or a drift ring.

2.1. Numerical methods

In general, there are two basic methods of calculating the LCEs, the derivative method [9] and the "pull-back" method [66]. The derivative method requires integrating or iterating the linear part of the derivative forward in time, thus necessitating the integration or iteration of d^2 equations to calculate all the exponents. The "pull-back" method, on the other hand, requires integrating or iterating a perturbed trajectory and its derivative matrix for each exponent desired; this can become very computationally costly — on the order of d^3 to compute all the exponents (see [26] for a comparison between various methods). Both methods require approximately the same numerical tools — an algorithm used to determine the singular values of the derivative matrix which are then renormalized. Efficient algorithms for doing such can be found in [30] or [54], and the numerical accuracy of such algorithms can be found in [31]. A thorough numerical analysis of these methods has only just begun (see [18,19], or [45]). However, these point toward the LCE algorithms as being highly stable in the sense that each of the individual steps is numerically stable. However, the sequence of these combined algorithms has not yet undergone careful study. The computer code used to calculate the LCE spectrum for our neural networks has been benchmarked with many known cases, performing well in each case. The program we use to calculate the LLE using the "pull-back" method has been similarly benchmarked. For the dynamical systems we consider, the LLE and LCE algorithms agree on the magnitude of the largest exponent up to statistical errors.

3. Universal approximators

Scalar neural networks like the ones we will use in this paper are "universal approximators", meaning they can approximate any mapping from many very general spaces of mappings if given enough neurons (e.g., any C^r mapping and its derivatives on compacta, varieties of Sobolev spaces). That scalar neural networks can approximate the mappings of interest is a topic addressed in detail in a paper by Hornik et al. [33]. Scalar neural networks such as those discussed in Ref. [33] form a class of time-delay dynamical systems. The motivation for the selection of these neural networks is two-fold. First, a combination of the approximation theorems of Hornik et al. [33] and the time-series embedding results in Takens' embedding theorem [65], or the embedology of Sauer et al. [60], shows this type of network is equivalent[3] (here the notion of equivalence is an embedding of C^r manifolds) to dynamical systems mapping compact sets in R^n to themselves and can yet accept a measure (for the specific construction, see [2]). Second, neural networks are a practical space of mappings used by time-series analysts to reconstruct unknown dynamics from time-series data [27,62, 37].

3.1. Neural networks

The discrete-time mappings we will consider are single-layer, feedforward neural networks of the form

$$x_t = \beta_0 + \sum_{i=1}^{n} \beta_i \tanh\left(s\omega_{i0} + s\sum_{j=1}^{d} \omega_{ij} x_{t-j}\right) \tag{1}$$

which is a map from R^d to R; the class of neural networks is denoted $\sum(\tanh)$. In Eq. (1), n represents the number of hidden units or neurons, d is the input or embedding dimension of the dynamical system which functions simply as the number of time lags, and s is a scaling factor on the weights. The parameters are chosen in the following way: $\beta_i, w_{ij}, x_j, s \in R$, where the β_i's and w_{ij}'s are elements of weight matrices (which we hold fixed for each case), $(x_1, \ldots, x_d)$ represents the initial condition, and $(x_t, x_{t+1}, \ldots, x_{t+d-1})$ represents the current state of the system at time t. We assume that the β's are *iid* uniform over [0, 1] and then re-scaled to satisfy $\sum_{i=1}^{n} \beta_i^2 = n$. The w_{ij}'s are *iid* normal with zero mean and unit variance (i.e., $N(0, 1)$). The s parameter is a real number and can be interpreted as the standard deviation of the w matrix of weights. The initial x_j's are chosen *iid* uniform on the interval $[-1, 1]$.

For x very near 0, the $\tanh(x)$ function is nearly linear; for $|x| \gg 1$, $\tanh(x)$ will tend to behave much like a binary function (± 1); for $|x|$ at intermediate values $\tanh(x)$ behaves as a highly nonlinear function. Thus, the scaling parameter s provides a unique bifurcation parameter that sweeps from linear ranges to highly non-linear ranges, to nearly binary ranges — fixed points to chaos to (finite-state) periodic orbits. This paper addresses the s-interval solely between the first bifurcation and the onset of chaos. The dynamics up to the first bifurcation are presented in [1] and [5]; the dynamics after the onset of chaos are discussed in [4,2] and [3].

3.1.1. The measure on $\sum(\tanh)$

We impose a measure on the space of neural networks via the probability distributions of the space of parameters, R^p where $p = n(d + 2) + 2$. Thus each neural network can be identified with its parameter values and thus a point in R^p — so the measure on the parameters is a measure on $\sum(\tanh)$. We will denote the individual measures on the parameters with $m_{\omega_{ij}}$, m_{β_i}, and m_{IC} for the ω's, β's, and the initial conditions respectively. As previously specified, $m_{\omega_{ij}} = N(0, s)$ for all i and j, and m_{β_i} is uniform on [0, 1] for all i. This forms a product measure on R^p. Understanding how this measure affects the observed dynamics is a major point of this paper; however, it is worth making a few initial points here. First, all the mappings in the class of neural networks we are studying are included (up to a renormalization factor) with the product measure we employ. However, because the measure determines what we observe, not all phenomena will occur with equal likelihood. Second, viewed from the perspective of reconstruction, the space $\sum(\tanh)$ includes delay coordinates as necessary to allow for the reconstruction of a dynamical system. Because all of the time-delay coordinates are included via the same measures, the past and present coordinates carry equal weight — thus all time and space scales are evenly represented and mixed. This feature is fairly well represented in the LCE spectrum after the onset of chaos. However, understanding how, and in what ways, the uniformity of measures across past states affects the dynamics along the route to chaos will require comparisons with similar studies having different imposed measures. Third, β_0 defines the symmetry about zero ($\beta_0 = 0$ forces a symmetry about zero); and as n is increased, symmetry about zero becomes less likely. However, aside from the particular bifurcations (e.g., saddle node versus pitchfork), the symmetry about zero seems to have little effect on the conclusions of this work. Finally, endowing the ω matrix with a distribution with significant non-zero mean forces the neurons to saturate, allowing for only periodic dynamics and thus no route to chaos.

The weight structure we employ involves parameters that are uncorrelated; this provides the simplest construction for our initial study. This choice implies that space and time scales are being combined via uncorrelated parameters — how this affects the observed dynamics is a major thrust of this work. Before a full understanding of the implications of the construction can be made, other similar studies must be conducted with different imposed measures. It is likely — though not known — that other probability measures using *iid* weights will generate similar results. The most exciting comparisons will likely lie with correlated weight structures. Correlating parameters yields a joint probability distributions on R^p instead of a product measure structure. Such situations are present in neural networks that undergo training on a particular model; the correspondence between various physical systems and neural

[3] For a clear understanding of the restrictions specifying when time-delay dynamical systems are equivalent to dynamical systems without a time-delay, see [60,49].

networks with uncorrelated versus correlated weight structures is an open question. Moreover, despite the fact that product and joint probability distributions are fundamentally different types of distributions, the effects of such on the space of neural networks is completely unknown. As discussed in Section 1.2, it is hoped that notions allowing for comparison in measures will stimulate a dialog between more abstract, computational dynamics and natural science communities.

4. Random matrix theory and the route to chaos

Doyon et al. [20] argued that, as the dimension of the dynamical system is increased, the most likely route to chaos in their set of dynamical systems would be via the Ruelle–Takens route to chaos. Their arguments were largely based on the random matrix results of Girko [28] (for other useful versions of this random matrix result, see Edelman [23] and Bai [8]). More precisely, the claim is that as the dimension of a dynamical system is increased, any bifurcations due to real eigenvalues (e.g., flip or fold bifurcations) will be vanishingly rare; and the route to chaos from a fixed point in parameter space will consist of periodic orbits with high-period (>4) and quasi-periodic orbits. The largest difference between the conclusions of Doyon et al. and the claims of Ruelle and Takens is that there may be more than two or three bifurcations before the onset of chaos.

A basic sketch of the matrix theory result of Girko [28], Edelman [23], and Bai [8] that is relevant to the current scenario is the following: given a square matrix whose elements are real random variables drawn from a distribution with a finite sixth moment, in the limit of infinite dimensions, the normalized spectrum (or eigenvalues) of the matrix will converge to a uniform distribution on the unit disk in the complex plane. It is worth noting that the convergence in measure with increasing dimension is not absolutely continuous with respect to Lebesgue measure. Nevertheless, if the Jacobian of the map at the "first" bifurcation point (i.e., the bifurcation from the fixed point) is a high-dimensional matrix whose elements are from a probability distribution with a finite sixth moment,[4] it is reasonable that the bifurcation would be of type Neimark–Sacker (via complex eigenvalue with irrational angle) instead of a flip or fold (via a real eigenvalue), with probability approaching unity as the dimension goes to infinity. For dynamical systems whose linear part is generated from a random distribution with finite sixth moment, it is not difficult to show (c.f. [1]) that the probability of the first bifurcation (via a linear bifurcation parameter) being of type Neimark–Sacker approaches unity. Extending this analytically for the kth bifurcation before the onset of chaos, where $k > 1$, is largely a matter of reducing the bifurcations of quasi-periodic orbits to bifurcations of fixed points. The general scheme for using random matrix theory to prove probable routes to chaos begins using the linear portions of all the known versal deformations (generalized normal forms [41,7]) which will correspond to a matrix, and then using random matrix results to show that the most probable bifurcation will be *to* a quasi-periodic orbit or an orbit with period $q > 4$. This research program is currently hindered by the fact that the normal form and center manifold theory for quasi-periodic bifurcation theory is far from complete (e.g., the codimension 2 situation is not complete yet, and the codimension 3 case is even further from completion). However, assuming that, in the end, most bifurcations of periodic and quasi-periodic orbits can be captured by some sort of Taylor series expansion (via a vector field approximation or a suspension); then, even though the linear term of the Taylor expansion will be degenerate and the outcome of the bifurcation will be determined by contributions of higher-order terms, the degeneracies in the linear term of the Taylor polynomial will nevertheless be due to complex eigenvalues — leading to some sort of a bifurcation (yet to be understood) from quasi-periodic or periodic orbits to other quasi-periodic or periodic orbits.

There are clearly portions of the above intuition that are not always correct or do not apply in all circumstances. For instance, in [1] it is demonstrated that for a random Gaussian matrix, the largest real eigenvalue scales with the mean while the largest complex eigenvalue remains unchanged. Thus, if the Gaussian distribution has a large mean, complex bifurcations will never be observed. Likewise, assuming the elements of the derivative are *iid* is not always a sound assumption. However, as the point of this random matrix discussion is merely as an intuitive framework, we will leave investigation of these issues for other studies.

4.1. Random matrix theory and time-delay dynamical systems

Elements of the space of dynamical systems we will be employing, time-delay neural networks, like all time-delay dynamical systems, have derivatives of a special type. In particular, Df, the linear part of the derivative at a point, is a companion matrix. Thus, for the first bifurcation, random polynomials are of particular interest instead of general random matrix theory (the spectrum of random companion matrices can be identified with the spectrum of polynomials with random coefficients). We have investigated the first bifurcation beginning in [5] and in considerably greater detail in [1] using results of Edelman et al. [24]. Results linking distributions of the coefficients of random polynomials to the distributions of the spectra of the respective polynomials are considerably more diverse than the analogous results for (full) random matrices. Thus, the first bifurcation results for networks such as ours are somewhat more complicated than for dynamical systems described in the previous section. However, in this work, we are concerned with the route to chaos or the sequence of bifurcations rather than the first bifurcation. In such a circumstance, the Jacobian is a product of the tangent map

[4] Nearly all probability distributions have a finite sixth moment — the Levy probability distribution is a notable example for which the sixth moment is not finite. An important probability distribution of this type is the Gaussian normal distribution for which Girko first proved his now famous "circular law". Bai generalized (as did Girko) and clarified this result to include all probability distributions with a finite sixth moment. Edelman arrived at similar conclusions via numerical analysis and proved various convergence rates to the circular law.

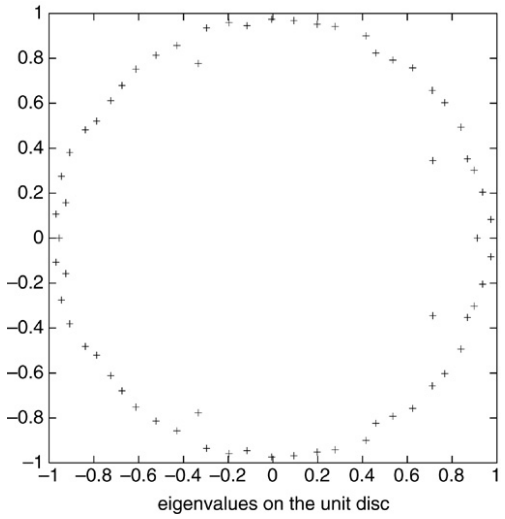

Fig. 1. The spectrum of the example network that will be analyzed in Section 5 for $s = 0.01$ — before the first bifurcation.

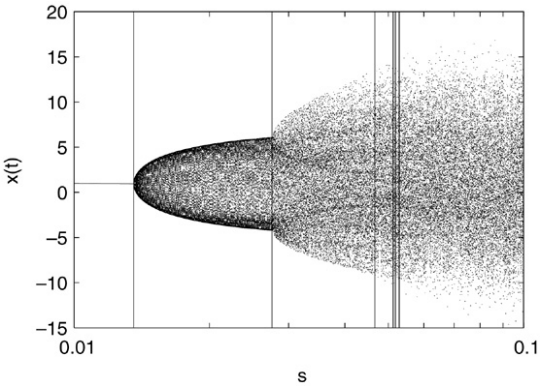

Fig. 2. Bifurcation diagram for a typical network with $n = 32$, $d = 64$. For each s value the initial conditions were reset and 100,000 time-steps were iterated before plotting the data (which consists of 100 points per s value). The vertical lines are s values of six bifurcation points and correspond to s values of 0.0135, 0.02755, 0.04667, 0.05124, 0.05183, and 0.05289 respectively.

along the orbit, and products of random companion matrices are often full random matrices. Showing relationships between distributions of spectra of particular matrices and the resulting spectra of products of matrices is well beyond the scope of this paper. Using results from [24], it might be possible to calculate the distribution of eigenvalues for neural networks at fixed points with the measures we have imposed on the parameter space. We have not carried out this calculation; instead we have calculated the spectrum for a particular example that will be carefully analyzed in Section 5 at an s value where the neural network is at a fixed point. Considering Fig. 1 (the spectrum of the example neural network before the first bifurcation), and the fact that increases in s will correspond to linear increases in the modulus of the eigenvalues, one can see why bifurcations from fixed points due to real eigenvalues for the dynamical systems we study will be less likely as the dimension is increased (i.e., the fraction of the eigenvalues that are real decreases monotonically with dimension). However, because the normal form theory is yet incomplete, we are restricted to using random matrix results as an intuitive framework and not a rigorous one when discussing sequences of bifurcations.

5. Numerical analysis: Prototypical example

Due to the diversity of the specifics of the networks we have studied, we consider networks individually with four types of figures: a standard bifurcation diagram; the largest Lyapunov exponent computed independently of the full spectrum; phase-space diagrams; and the full Lyapunov spectrum.

Our choice of the number of neurons and the number of dimensions is based on Fig. 1 of [5] and the compromise required by computational time limits. Considering Fig. 1 of [5], 32 neurons puts our networks deeply in the region of the set of neural networks that correspond to highly complicated and chaotic dynamics. The dimension of 64 was chosen because it was the lowest dimension that is highly representative of $d > 64$. The compute time increases as a power of the dimension, and thus we are required to strike a balance between computation time and the number of cases we could consider. The case we will consider is typical of the 500 networks we studied with $d = 64$. We will draw attention to one feature that is not common as n and d are increased. We have included this feature to demonstrate the diversity of networks with intermediate numbers of parameters and dimensions. Aside from this feature, there is very little difference between the networks with $n = 32$ and $d = 64$ and most networks with much higher n and d values, provided the networks are restricted to the parameter region that corresponds to the route to chaos. Increasing d lowers the s value where this route occurs. Increasing the number of neurons has little effect in this particular region aside from decreasing the existence of low-period periodic orbits. The n dependence will be discussed in detail in a later section where we address Monte Carlo results. Increasing n does have a profound effect for higher s values in the chaotic portion of parameter space because the number of neurons controls the entropy of the network (see [44] or [3] for more details). Lowering the number of neurons can also have a large effect in the sense that if d is increased and n is fixed, the entropy appears to decrease to zero — thus, eventually there is no route to chaos. However, these low-n networks are of little interest to our present direction of study because they afford a more crude representation of C^r function space.

5.1. Bifurcation diagrams

Beginning with Fig. 2, the standard bifurcation diagram, there are four important features to notice. The first feature is the first bifurcation which occurs at $s = 0.0135$ from a fixed point to a drift ring (invariant circle). A secondary bifurcation is clearly visible at $s = 0.02755$, the nature of this bifurcation is unclear from the perspective of Fig. 2. Chaos seems to onset near $s = 0.05$, and has surely onset at $s = 0.06$. However, the exact location of the onset of chaos is difficult to discern from this figure. Lastly, all of the dynamics between the fixed point and chaos consist of n-tori for $n \geq 0$.

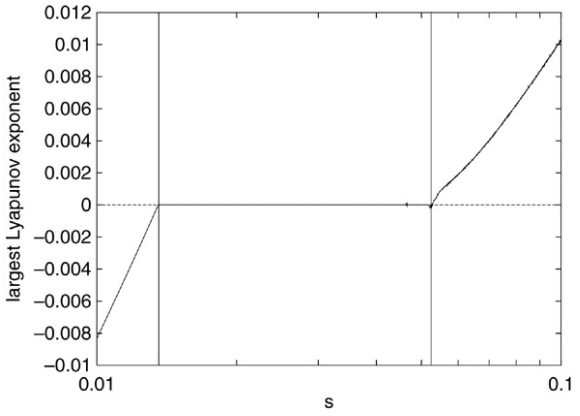

Fig. 3. The largest Lyapunov exponent for a typical network; $n = 32$, $d = 64$. For each s value the initial conditions were reset and 10^5 time-steps were iterated before calculating the LLE. The LLE was calculated over the following 50,000 time-steps. The vertical lines are s values of the first and last bifurcation points before chaos and correspond to s values of 0.0135 and 0.05289 respectively.

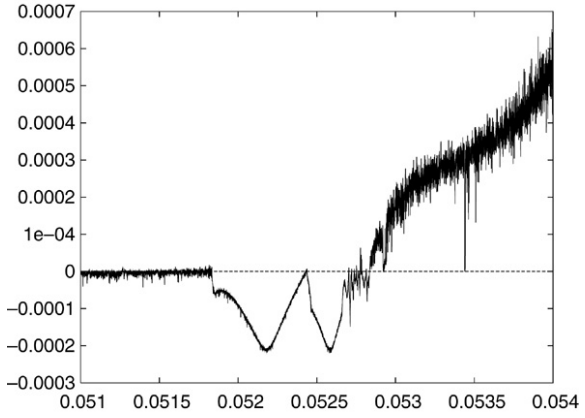

Fig. 4. The largest Lyapunov exponent at the onset of chaos. For each s value the initial conditions were reset and 10^5 time-steps were iterated before calculating the LLE. The LLE was calculated over the following 2×10^6 time-steps.

5.2. Largest Lyapunov exponent

Next, consider Fig. 3 — the largest Lyapunov exponent versus s. Again, as expected, we see the first bifurcation at $s = 0.0135$, in agreement with the bifurcation diagram. Fig. 3, however, gives a clear picture of the onset of chaos which occurs at $s = 0.05289$. The largest exponent is near zero between the first bifurcation and the onset of chaos providing evidence for the existence of at least one persistent eigenvalue(s) with modulus one (assuming a Jacobian can be constructed).

Considering Fig. 3, near the onset of chaos the exponent becomes negative over a very short s interval. Ignoring all the intermediate bifurcations, consider the onset of chaos via Fig. 4. In this figure there is evidence of the existence of an apparent periodic orbit followed by likely a period doubling bifurcation that is in turn followed by a complicated bifurcation structure. Besides noting this for general interest and completeness, we will refrain from a further discussion of this small interval since this behavior seems to disappear for high-dimensional networks and is not particularly related to the point of this paper.[5]

5.3. Phase portraits

We will begin our presentation of phase-space figures at the second bifurcation. Considering Fig. 2, the second bifurcation corresponds to the rapid rate of change in the attractor size near $s \sim 0.0275$. The nature of these bifurcations cannot be fully determined by a consideration of the Lyapunov spectrum or the largest Lyapunov exponent. This is why it is important to consider the phase space diagrams. The second bifurcation appears to be from a 1-torus (drift ring) to a 2-torus as shown in Fig. 5.

The phase-space plots on either side of the third bifurcation, which is not clearly apparent in the bifurcation diagram, the LLE plot, or in the Lyapunov spectrum (as we will see) is depicted in Fig. 6. Just before the third bifurcation the smooth looking torus from Fig. 5 has become "kinked" significantly. At the third bifurcation, the torus-like, two-dimensional object becomes a one-dimensional object. Thus the third bifurcation is from a 2-torus to a 1-torus and occurs at $s \sim 0.0466$. The 1-torus is a severely "kinked" quasi-periodic orbit.

Considering the phase-space plots on either side of the fourth bifurcation in Fig. 7 one might conclude that our example has undergone a period doubling of the quasi-periodic orbit. An analytical explanation of such a bifurcation is yet an open problem, but it is likely that this bifurcation is of "Neimark–Sacker-Flip" type. We will refrain from a further discussion of this bifurcation, directing the interested reader to chapter 9 of [41] for more information.

We do not present figures for the fifth bifurcation, and simply note that it is a bifurcation from a quasi-periodic orbit on the 1-torus to a high-period, cyclic orbit. Fig. 8 illustrates the sixth and final bifurcation into chaos. However, in our particular example, considering Fig. 4, just before the onset of chaos, there is likely a sequence of bifurcations. We will not belabor this further because little insight is gained from further consideration, and the inclusion of this case is only meant to serve as a demonstration of the diversity that is encountered in the specific mappings that are presented in this paper. This sequence of bifurcations involving a period-doubling just before the onset of chaos is increasingly rare as the dimension increases.

5.4. Lyapunov spectrum

Using the LCE spectrum to locate bifurcation points prior to the appearance of a positive LCE is achieved by considering the number of exponents that are zero — at points where the number of exponents equal to zero changes, there was a bifurcation point. The first bifurcation in Fig. 9 occurs at $s = 0.0135$ and signals the Neimark–Sacker type first bifurcation,

[5] This is the slightly atypical behavior we alluded to in the beginning of Section 5.

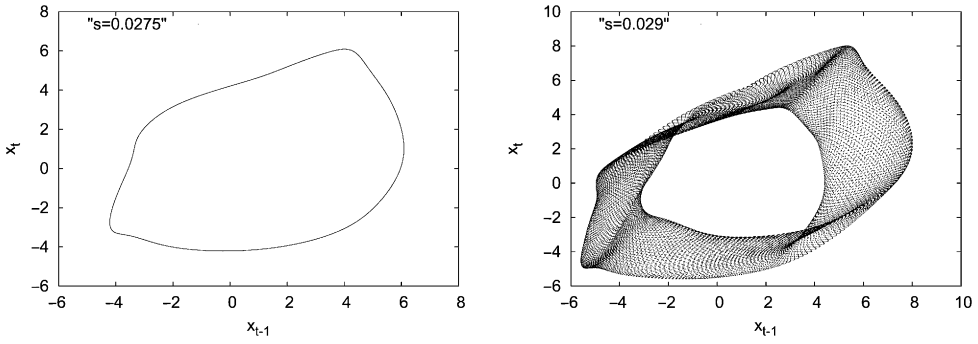

Fig. 5. Phase plots on either side of the 2nd bifurcation, $s = 0.0275$ and $s = 0.029$ respectively. The bifurcation occurs at $s \sim 0.02754$. The network was iterated for 10^6 time-steps before the 30,000 time-steps were kept for this plot.

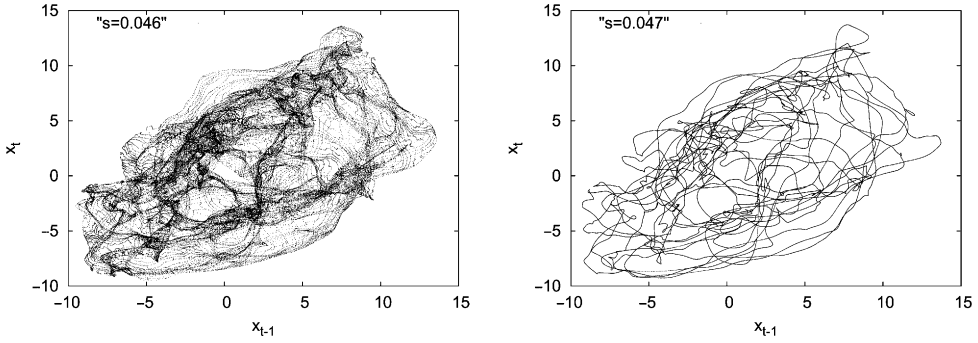

Fig. 6. Phase plots on either side of the 3rd bifurcation, $s = 0.046$ and $s = 0.047$ respectively. The bifurcation occurs at $s \sim 0.04667$. The network was iterated for 10^6 time-steps before the 30,000 time-steps were kept for this plot.

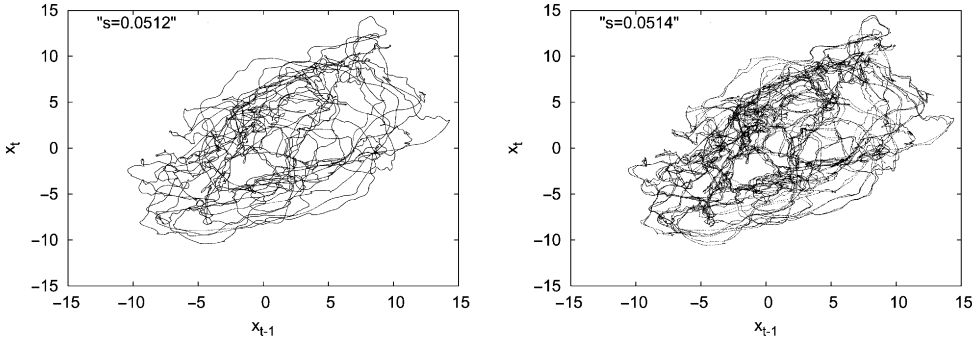

Fig. 7. Phase plots on either side of the 4th bifurcation, $s = 0.0512$ and $s = 0.0514$ respectively. The bifurcation occurs at $s \sim 0.05124$. The network was iterated for 10^6 time-steps before the 30,000 time-steps were kept for this plot.

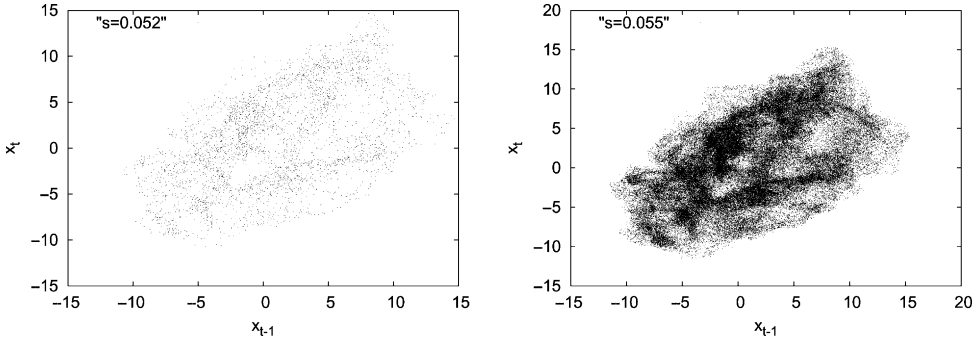

Fig. 8. Phase space plot near the 6th bifurcation, $s = 0.052$ and $s = 0.055$ respectively. The bifurcation occurs at $s \sim 0.05294$. The network was iterated for 10^6 time-steps before the 30,000 time-steps were kept for this plot.

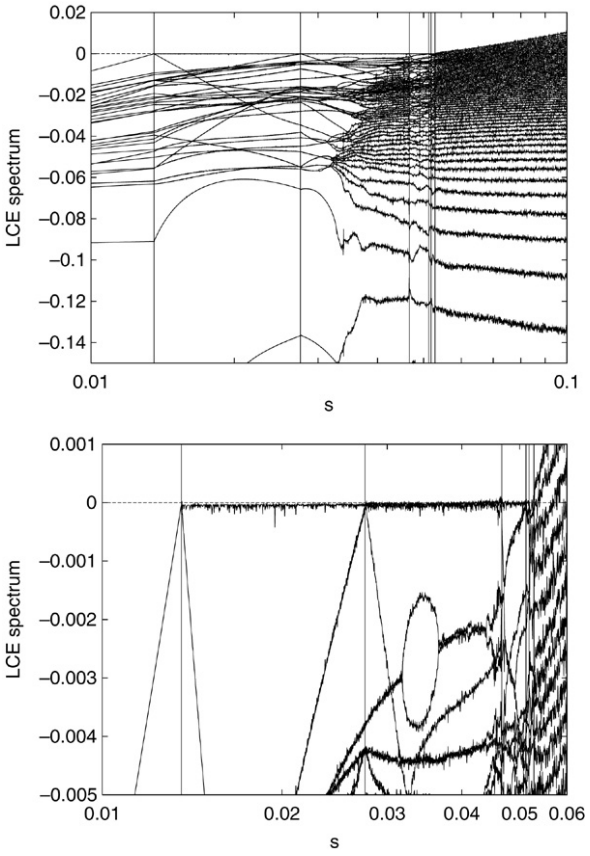

Fig. 9. Lyapunov spectrum of the typical network; $n = 32$, $d = 64$. The s parameter was incremented by the factor 1.001. For each s value the initial conditions were reset and 10^5 time-steps were iterated before calculating the LCEs; the LCEs were then calculated over the following 40,000 time-steps. The vertical lines are s values of six bifurcation points and correspond to s values of 0.0135, 0.02754, 0.04667, 0.05124, 0.05183 and 0.05289 respectively. Both plots (top and bottom) are of the same network — with different scales to depict different relevant features.

as expected. The second bifurcation occurs at $s = 0.02755$ and consists of an exponent pair touching zero. After the bifurcation point, one of these exponents remains zero and the other decreases below zero, indicating a bifurcation to a 2-torus. The third bifurcation can been observed to occur at $s = 0.04667$ where one of the two zero exponents decreases below zero, indicating a bifurcation to a limit-cycle. The remaining bifurcations up to and including the bifurcation to chaos are identical to those seen in the section discussing the largest Lyapunov exponent.

The low-s portion of Fig. 9 is surprising because there are only about half as many Lyapunov exponents as dimensions. This is because, on the s interval $\sim (0, 0.03)$, nearly all the exponents correspond to pairs of complex eigenvalues. At $s \sim 0.03$ the most negative exponents begin to separate into distinct values. This cascade continues to occur from the most negative exponents to the largest exponents, in that order, up to the onset of chaos. For s at the onset of chaos, there exist 64 distinct exponents, one for each dimension. In fact, by $s \sim 0.052$, all the exponent pairs are observably split. After the onset of chaos, in a very short s interval, the high-dimensional neural networks undergo a transition to high-entropy systems [2]. Our interpretation of this transition from fixed points to a high-entropy system in our (dissipative) set of dynamical systems occurs by the gradual degeneration or the tearing apart of the stable torus. Before the first bifurcation, all or nearly all of the directions correspond to stable sinks with complex eigenvalues. Taking the Cartesian product of these sinks yields a $\frac{d}{2}$-torus (we are assuming all the exponents are complex for simplicity) with a stable fixed point. This torus will only be observed if the transient dynamics are observed. Once the transients die, only a fixed point will remain. If the fixed points were not due to complex eigenvalues, the transient dynamics would not occur on a torus but rather on a collection of lines. When a Neimark–Sacker bifurcation occurs, one of the dimensions on the $\frac{d}{2}$-torus becomes a drift ring, one becomes a sink due to the real eigenvalue, and the rest remain sinks due to complex eigenvalues. As the dynamical system makes the transition to chaos, the dimensions that are nearly neutral undergo various bifurcations as we have observed. However, starting with the more strongly contracting directions which are not easily observed, the torus that corresponds to the transient dynamics begins to tear apart. The stable rotating dimensions separate in pairs, tearing the contracting directions of the torus apart. What is left just before the onset of chaos are neutral and very nearly neutral directions that have rotation and strong stable directions (sinks) with no coherent rotation — the $\frac{d}{2}$-torus is gone, and all that remains is the observable m-torus with $m \leq 3$. Then, the neutrally stable directions tear apart, corresponding to the breakup of the observable limit-cycle or torus and the onset of chaos. This is a geometric interpretation of the transition to chaos in the high-dimensional dynamical systems we study in this paper in accordance with the Lyapunov spectrum.

Once the onset of chaos has occurred, we can observe the contrast between the Ruelle–Takens version of chaos, which consists of strange attractors, and the Landau and Hopf model, which consists of interacting quasi-periodic orbits. Strange attractors are not a collection of interacting quasi-periodic orbits or a rotating soup, but rather, distinct directions of expansion, contraction *à la* axiom A, and a little bit of rotation (neutral directions). The existence of a distinct exponent for each direction in the chaotic region displays this difference nicely. The cascade from all commensurate frequencies before the onset of chaos is observed to all incommensurate Lyapunov exponents after chaos is observed was unexpected.

5.5. Summary of the prototypical case

Table 1 contains a summary of the information that can be gleaned from the bifurcation diagram, the LLE, the phase plots and the LCE spectrum; this is information regarding the location of the bifurcations and the dynamics between the first bifurcation from a fixed point to the appearance of chaos for this neural network. Table 1 distills the combination of this information into a unified picture.

Table 1
Summary of the route to chaos for the prototypical network

Bif. number	s at the 1st	s at the 2nd	s at the 3rd	s at the 4th	s at the 5th	s at the 6th
Largest Lyapunov exponent	0.0135	–	–	–	0.05183	0.05289
Bif. diagram	0.0135	0.02755	–	–	–	0.0525
Phase space diagram	0.0135	0.02754	0.04667	0.05124	0.05183	0.05289
Lyapunov spectrum	0.0135	0.02755	0.04667	0.05124	0.05183	0.053
Transition to:	T^1	T^2	T^1	T^1	T^0	Chaos

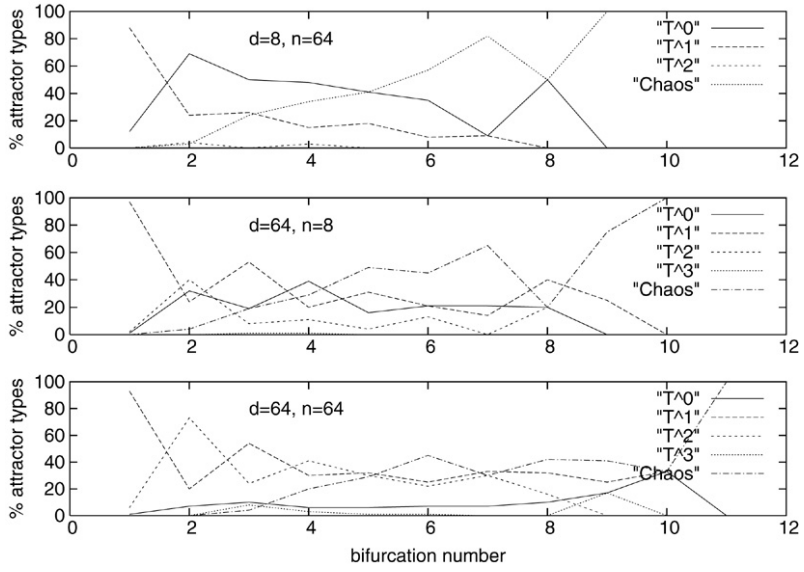

Fig. 10. The percentage of attractor types following each bifurcation for ensembles of neural networks with $n = 64$ and $d = 8$, $n = 8$ and $d = 64$, and $n = 64$ and $d = 64$. For each n and d pair at least 200 networks were considered. Once networks became chaotic, further bifurcations were no longer counted.

6. Statistical analysis

Summarizing the likelihoods of bifurcation sequences for the dynamical systems becomes computationally unmanageable as the dimension of the dynamical system increases because the number of bifurcations along the route to chaos increases with increasing dimension. Even with bifurcations between only five attractor types and as many as 11 bifurcations per mapping, the number of possible sequences is in the hundreds of thousands; a proper statistical study of bifurcation sequence frequency could require considering millions of networks and hundreds of thousands of sequences of bifurcations. Instead of considering all bifurcation sequences explicitly, we will consider, for a variety of n and d values, the single most likely bifurcation sequence to chaos and the percentage of attractor types after each bifurcation.

The most common sequence of bifurcations for the ensembles of networks used to generate Fig. 10 are, coincidentally, the same sequences that could be deduced by a consideration of the percentages given in Fig. 10. These sequences are:

$n = 64, d = 8$: $T^1 \to T^0 \to T^0 \to T^0 \to$ Chaos
$n = 8, d = 64$: $T^1 \to T^2 \to T^1 \to T^0 \to$ Chaos
$n = 64, d = 64$: $T^1 \to T^2 \to T^1 \to T^2 \to T^1 \to$ Chaos.

In general one must be careful about what is inferred about probable bifurcation sequences from Fig. 10. In the case where $d = 64$ and $n = 8$, the second most common first bifurcation was to a T^2 attractor. From the data in Fig. 10, given a first bifurcation to T^2, it is not possible to deduce what the most probable second bifurcation and subsequent attractor type will be. Nevertheless, Fig. 10 can provide significant information regarding the sequence of bifurcations en route to a chaotic attractor.

Increases in n and d have some similar effects on the dynamics and bifurcations, as well as on some non-overlapping effects. In the high-dimensional networks, increasing n increases the likelihood of T^3 attractors. Moreover, increasing n simultaneously decreases the likelihood of bifurcations to T^0 attractors while increasing the existence of T^1 and T^2 attractors. In low-dimensional networks, increases in n increase the meager likelihood of observation of T^2 attractors. Increases in d have more profound effects. Increases in d greatly increase the probability of observing T^2 attractors while greatly reducing the probability of observing T^0 attractors. In low-dimensional networks, T^3 attractors are never observed; thus, increases in the dimension allow for the existence of stable, observable T^3 attractors. The high-d, low-n networks have the most complicated bifurcation sequences as they have the most diverse attractor types. The effect of increasing n and

d is most concisely summarized by recalling that the most common sequence of bifurcations for the $n = 64$, $d = 64$ ensemble of networks is a sequence of T^1 and T^2 attractors leading to a chaotic attractor. All of these results lead to the conclusion that for the construction we use in this paper, increasing n and d will lead to a route to chaos that consists of bifurcations between T^1 and T^2 attractors. It is important to note that increasing the dimension by four octaves produced no observed T^4 attractors. This does not mean T^4 attractors do not exist in our construction (clearly for some measures on the parameter space they must), or even that they will not be observed when d is much larger than we consider in this work. However, since T^3 attractors are observable for $d \geq 16$ (with differing likelihoods), whereas T^4 attractors are not, it is possible that observing T^k attractors for $k > 3$, is unlikely in our construction for any d. Nevertheless, we are not willing to claim that there exists an N such that as $d \to \infty$, $k < N$ for T^k, let alone $N = 4$. Finally, even if there does exist an upper bound on the dimension of the observable tori in our experiment, one of the key features of the framework we are employing is the ability to quantify, relative to a *measure*, observable dynamics. Certainly other measures exist such that T^d attractors exist; however, it is likely that this measure is not very representative of many observable phenomena in nature. Addressing this last issue is a source of future study.

7. Summary

The example we presented gives the flavor both of the commonality and diversity that exists along the route to chaos in the ensemble of networks we studied. For the high-d, high-n networks with which we are most concerned in this paper, the predominant route to chaos consisted of bifurcations between T^1 and T^2 attractors. However, there is a lot of variation both in the specific sequences of bifurcations and in the number of bifurcations between the first bifurcation and the onset of chaos. Period-doubling cascades and such routes are exceedingly rare — we did not observe a period-doubling route to chaos in the 2500+[6] high-d networks we considered for this paper. Furthermore, in any of the 10000+ networks we studied, we did not observe a period-doubling sequence leading to chaos, akin to what is observed in the 1-dimensional logistic map. While there do exist sequences of bifurcations between periodic orbits in low dimensions, we never observed long sequences of real bifurcations in high-dimensional cases. Moreover, in the low-dimensional cases, many of the bifurcations from between periodic orbits were not period doubling bifurcations, but rather bifurcations between high-period, periodic orbits. Nevertheless, we did observe a substantial number of sequences of period doubling bifurcations in low-d networks. This is largely in agreement with the work of Doyon et al., Cessac et al., and Sompolinsky et al. despite the fact that they study a considerably different space of dynamical systems. We do find differences of course — the routes to chaos interval in parameter space occupies roughly two (ever decreasing) octaves — but the differences are often superficial. In fact, considering the Euclidean length (in parameter space) of the route to chaos for the neural networks we study is somewhat deceptive because increasing the number of terms (e.g. the number of neurons) in the argument of tanh effectively increases the magnitude of the argument of tanh, which lowers the threshold where the linear regime is breached, where the onset of chaos occurs, and where the onset of finite-state dynamics begins. Consideration of the length of the routes to chaos region is better done in terms of decades of the parameter, largely because $s_{\text{first bifurcation}} \sim n^a d^b < s_{\text{chaos onset}} \sim n^{\tilde{a}} d^{\tilde{b}}$ where $|a| < |\tilde{a}|$ and $|b| < |\tilde{b}|$ at respective n and d values. Moreover, the quantity $|\log(s_{\text{chaos onset}}) - \log(s_{\text{first bifurcation}})|$ increases when n and d are simultaneously increased. There is likely a deep explanation related to random matrix theory for the similarity between the results in this paper and the results of Doyon et al., Cessac et al., and Sompolinsky et al.; however, the details are yet unresolved. Finally, we rarely observe 3-tori. Thus, the results in this paper are consistent with the prediction of Newhouse, Ruelle and Takens [47] that would claim that 3-tori would be uncommon because they are easily perturbed to strange attractors. However, often when we did observe 3-tori, they persisted for relatively long s-intervals.

The primary focus of this paper is the bifurcation sequences along the route to chaos. However, there is a feature of the evolution of the attractors along the route to chaos that, while not appearing in the portion of the attractor with positive measure, is still highly relevant to the transition to chaos. Before the first bifurcation, many of the strongly contracting directions consist of simple sinks with rotation due to complex eigenvalues coupling pairs of state-space dimensions. Leading up to the onset of chaos, the strongly contracting directions with rotation begin to decouple, beginning with the most strongly contracting directions. We believe that represents the following phenomena: the original fixed point lies on a torus; the decoupling of the stable directions corresponds to the tearing apart of this torus. Just before the onset of chaos, what remains is a T^0, T^1 or T^2 attractor where the rest of the stable directions correspond to sinks with no simple or coherent rotation. This "dissipative KAM theory" scenario is similar to what can be observed at the onset of chaos in Hamiltonian dynamical systems where the tori, as predicted by KAM theory, break up [39,40,7]. Fig. 11 shows the superimposed spectrum of the example neural network along the orbit at $s = 0.02$ where the attractor is T^1 with many stable "rotating" sinks and at $s = 0.05$ — just before the torus is torn apart and the onset of chaos occurs (yet only one LCE is zero). The attractor has become much more complex — the isolated sinks have decoupled but little can be said beyond this without a fixed point style analysis of the torus. This phenomena is far from being well understood.

One of the key reasons we chose the relatively complicated framework provided by the space of neural networks is the number of future pathways that are opened and the obstructions that are removed. The class of neural networks allows

[6] The high-d data set includes the 1000 $d = 64$ networks, 700 $d = 128$ networks, 700 $d = 256$ networks and networks of other miscellaneous dimensions up to $d = 1024$.

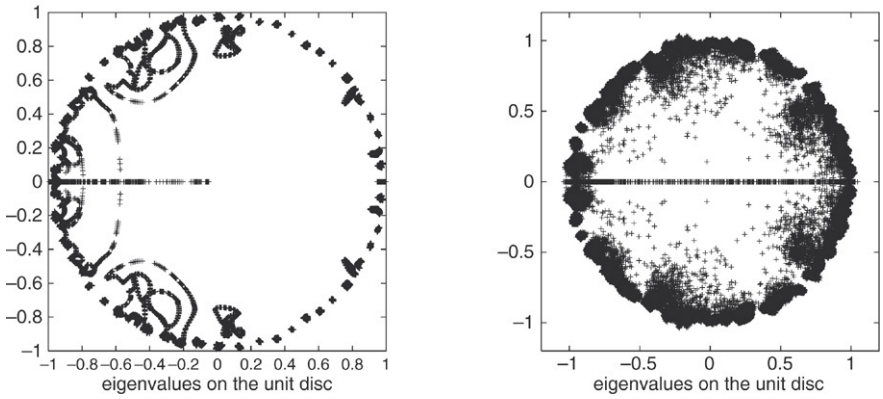

Fig. 11. The spectrum of the example network analyzed in Section 5 for $s = 0.02$ (left) — at a stable drift ring, and $s = 0.05$ (right) — just before the onset of chaos. This represents the spectrum *along* the orbit and not at a fixed point.

not only a practical (measure-theoretic) means of analyzing topological results, but it contains, in a manageable way, the supposed pathological examples. Due to the mappings that neural networks can approximate, if the spectrum of Lyapunov exponents of a d-dimensional network is contained in $[\chi_1, \chi_d]$, then it is likely there exists at least one path through parameter space such that any network can be transformed into a T^d torus with all Lyapunov exponents being zero. At this time we can only say this is likely because such a result has not been proven. If one were to stratify the networks by their spectra, the aforementioned torus would be but a point along the interval $[\chi_1, \chi_d]$, and thus, in this sense, unusual. Aside from the stratification of the space of neural networks with the LCE's, there exists the more standard stratifications of the networks, such as the map $\phi : R^{n(d+2)} \to \sum(\tanh())$, where $\sum(\tanh())$ is the set of neural networks with tanh() as the squashing function. Such a stratification allows for analysis akin to more standard bifurcation theory and the language of real varieties. Because of these available constructions, several problems are tempered by the neural network framework. For instance, the severe limitations of dynamical behaviors that exist in particular equations with few parameters can be solved by adding parameters without appreciably altering the model. Moreover, problems establishing measures or concrete spaces of parameters that exist with the study of C^r dynamical systems can be circumvented due to the existence of a flexible, concrete parameter space. Neural networks with few parameters behave like many of the concrete dynamical systems considered by scientists, and, as parameters are added, the neural networks become more free to behave like general C^r mappings that many mathematicians study. The neural network construction utilized here is a bridge between the more practical world of computational dynamical systems and the world of abstract dynamical systems.

Acknowledgments

We would like to thank R.A. Bayliss, W.D. Dechert, W. Brock, and Y. Sato for many helpful discussions and advice, and J.R. Albers, K. Connors, and J. Thompson for reading the manuscript. D.J. Albers would like to give special thanks to J.P. Crutchfield for guidance, many fruitful discussions, wonderful insight and feedback and support. Portions of the computing for the project were done on the Beowulf cluster at the Santa Fe Institute and was partially supported at the Santa Fe Institute under the Networks, Dynamics Program funded by the Intel Corporation under the Computation, Dynamics and Inference Program via SFI's core grant from the National Science and MacArthur Foundations. Computation was also performed on the Beowulf clusters at the Center for Computational Science and Engineering at the University of California - Davis and at the Max Planck Institute for Mathematics in the Sciences. Partial support for D.J. Albers was provided by NSF grants DMR-9820816 and PHY-9910217 and DARPA Agreement F30602-00-2-0583.

References

[1] D.J. Albers, J.C. Sprott, Probability of a local bifurcation type from a fixed point: A random matrix perspective, J. Stat. Phys. (in press). http://arxiv.org/abs/nlin.CD/0510060.
[2] D.J. Albers, J.C. Sprott, Structural stability and hyperbolicity violation in high-dimensional dynamical systems, Nonlinearity 19 (2006) 1801–1847.
[3] D.J. Albers, J.P. Crutchfield, J.C. Sprott, Phenomenological scaling in the organization of high-dimensional dynamics (in preparation).
[4] D.J. Albers, J.C. Sprott, J.P. Crutchfield, Persistent chaos in high dimensions, Phys. Rev. E (in press). http://arxiv.org/abs/nlin.CD/0504040.
[5] D.J. Albers, J.C. Sprott, W.D. Dechert, Routes to chaos in neural networks with random weights, Internat. J. Bifur. Chaos 8 (1998) 1463–1478.
[6] S. Amari, H. Nagaoka, S.-I. Amari, Methods of Information Geometry, in: Translations of Mathematical Monographs, AMS, 2001.
[7] V. Arnold, Geometric methods in the theory of ordinary differential equations, in: Grundlehren de mathematischen Wissenschaften, 2nd edition, Springer-Verlag, 1983.
[8] Z.D. Bai, Circular law, Ann. Probab. 25 (1997) 494–529.
[9] G. Benettin, L. Galgani, A. Giorgilli, J.-M. Strelcyn, Lyapunov characteristic exponents from smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application, Meccanica 15 (1979) 21–30.
[10] B. Braaksma, H. Broer, G Huitema, Toward a quasi-periodic bifurcation theory, Mem. Amer. Math. Soc. 83 (421) (1990) 83–175.
[11] H. Broer, Coupled Hopf-bifurcations: persistent examples of n-quasiperiodicity determined by families of 3-jets, in: Geometric Methods in Dynamics (I), in: Astérisque, vol. 286, 2003, pp. 223–229.
[12] H.W. Broer, G.B. Huitema, M.B. Sevryuk, Quasi-periodic Motions in Families of Dynamical Systems, in: Lecture Notes in Mathematics, vol. 1645, Springer-Verlag, 1996.

[13] H.W. Broer, G.B. Huitema, F. Takens, Unfoldings and bifurcations of quasi-periodic tori, Mem. Amer. Math. Soc. 83 (421) (1990) 1–81.
[14] K. Burns, D. Dolgopyat, Y. Pesin, Partial hyperbolicity, Lyapunov exponents and stable ergodicity, J. Statist. Phys. 108 (2002) 927–942.
[15] B. Cessac, B. Doyon, M. Quoy, M. Samuelides, Mean-field equations, bifurcation map and route to chaos in discrete time neural networks, Physica D 74 (1994) 24–44.
[16] H. Chaté, P. Manneville, Transition to turbulence via spatiotemporal intermittency, Phys. Rev. Lett. 58 (1987) 112–115.
[17] A. Chenciner, G. Iooss, Bifurcations de tores invariants, Arch. Ration. Mech. Anal. 69 (1979) 109–198.
[18] L. Dieci, R. Russell, E. Van Vleck, On the computation of Lyapunov exponents for continuous dynamical systems, SIAM J. Numer. Anal. 34 (1997) 402–423.
[19] L. Dieci, E. Van Vleck, Computation of a few Lyapunov exponents for continuous and discrete dynamical systems, Appl. Numer. Math. 17 (1995) 275–291.
[20] B. Doyon, B. Cessac, M. Quoy, M. Samuelides, Control of the transition to chaos in neural networks with random connectivity, Internat. J. Bifur. Chaos 3 (1993) 279–291.
[21] B. Doyon, B. Cessac, M. Quoy, M. Samuelides, On bifurcations and chaos in random neural networks, Acta Biotheo. 42 (1994) 215–225.
[22] J.-P. Eckmann, Roads to turbulence in dissipative dynamical systems, Rev. Modern Phys. 53 (1981) 643–654.
[23] A. Edelman, The probability that a random gaussian matrix has k real eigenvalues, related distributions, and the circular law, J. Multivariate Anal. 60 (1997) 203–232.
[24] A. Edelman, E. Kostlan, How many zeros of a random polynomial are real? Bull. Amer. Math. Soc. 32 (1995) 1–37.
[25] U. Frisch, Turbulence: The Legacy of A.N. Kolmogorov, Cambridge University Press, 1995.
[26] K. Geist, U. Parlitz, W. Lauterborn, Comparison of different methods for computing Lyapunov exponents, Progr. Theoret. Phys. 83 (5) (1990) 875–893.
[27] R. Gencay, W.D. Dechert, An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system, Physica D 59 (1992) 142–157.
[28] V. Girko, Circular law, Theory Probab. Appl. 29 (1984) 694–706.
[29] J.P. Gollub, H.L. Swinney, Onset of turbulence in a rotating fluid, Phys. Rev. Lett. 35 (1975) 927–930.
[30] G. Golub, C. Van Lorn, Matrix Computations, 3rd edition, The Johns Hopkins University Press, 1996.
[31] N.J. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, 1996.
[32] E. Hopf, A mathematical example displaying the features of turbulence, Comm. Pure Appl. Math. 1 (1948) 303–322.
[33] K. Hornik, M. Stinchocombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Netw. 3 (1990) 551.
[34] G. Iooss, M. Adelmeyer, Topics in Bifurcation Theory and Applications, World Scientific, 1998.
[35] G. Iooss, W.F. Langford, Conjectures on the Routes to Turbulence via Bifurcation, New York Academy of Sciences, 1980.
[36] G. Iooss, J.E. Los, Quasi-genericity of bifurcations to high dimensional invariant tori for maps, Comm. Math. Phys. 119 (1988) 453–500.
[37] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd edition, Cambridge University Press, 2003.
[38] A. Katok, Lyapunov exponents, entropy, and periodic orbits for diffeomorphisms, Publ. Math. Inst. Hautes Études Sci. 51 (1980) 137–174.
[39] T. Konishi, K. Kaneko, Diffusion in Hamiltonian chaos and its size dependence, J. Phys. A (1990).
[40] T. Konishi, K. Kaneko, Clustered motion in symplectic coupled map systems, J. Phys. A 25 (1992) 6283–6296.
[41] Y. Kuznetzov, Bifurcation Theory, 2nd edition, Springer-Verlag, 1998.
[42] D. Landau, L. Lifshitz, Fluid Mechanics, 1st edition, Pergamon Press, 1959.
[43] R. Mañé, A proof of the C^1 stability conjecture, Publ. Math. Inst. Hautes Études Sci. 66 (1988) 161–210.
[44] E.F. Manffra, Properties of systems with time-delayed feedback, Ph.D. Thesis, Univesitat Gesamthochschule Wuppertal, 2002.
[45] E.J. McDonald, D.J. Higham, Error analysis of the QR algorithms for computing Lyapunov exponents, Electron. Trans. Numer. Anal. 12 (2001) 234–251.
[46] J. Neimark, On some cases of periodic motions depending on parameters, Dokl. Acad. Nauk SSSR 129 (1959) 736–739.
[47] S. Newhouse, J. Palis, F. Takens, Bifurcation and stability of families of diffeomorphisms, IHES 57 (1983) 5–71.
[48] S. Newhouse, D. Ruelle, F. Takens, Occurrence of strange Axiom A attractors near quasiperiodic flows on T^m, $m \leq 3$, Comm. Math. Phys. 64 (1978) 35–40.
[49] W. Ott, J. Yorke, Learning about reality from observations, SIAM J. Appl. Dyn. Syst. 3 (2003) 297–322.
[50] W. Ott, J.A. Yorke, Prevalence, Bull. Amer. Math. Soc. 42 (2005) 263–290.
[51] J. Palis, Vector fields generate few diffeomorphisms, Bull. Amer. Math. Soc. 80 (1974) 503–505.
[52] M. Peixoto, Structural stability on two-dimensional manifolds, Topology 1 (1962) 101–120.
[53] Ya.B. Pesin, Lyapunov characteristic exponents and smooth ergodic theory, Russian Math. Surveys 32 (1977) 55–114 (English transl.).
[54] W.A. Press, S.A. Teukolsky, Portable random number generators, Comput. Phys. 6 (1992) 522–524.
[55] J. Robbin, A structural stability theorem, Ann. of Math. 94 (1971) 447–493.
[56] C. Robinson, Structural stability of C^1 diffeomorphisms, J. Differential Equations 22 (1976) 28–73.
[57] D. Ruelle, Characteristic exponents and invariant manifolds in Hilbert space, Ann. of Math. 115 (1982) 243–290.
[58] D. Ruelle, F. Takens, A route to turbulence, Comm. Math. Phys. 20 (1970) 167–192.
[59] R. Sacker, On invariant surfaces and bifurcation of periodic solutions of ordinary differential equations, Technical Report 333, New York State University, 1964.
[60] T. Sauer, J. Yorke, M. Casdagli, Embedology, J. Statist. Phys. 65 (1991) 579–616.
[61] H. Sompolinsky, A. Crisanti, H.J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. (1988).
[62] J.C. Sprott, Chaos and Time-series Analysis, Oxford University Press, 2003.
[63] H.L. Swinney, J.P. Gollub, The transition to turbulence, Phys. Today 31 (1978) 41–43, 46–49.
[64] H.L. Swinney, J.P. Gollub, Characterization of hydrodynamic strange attractors, Physica D 18 (1986) 448–454.
[65] F. Takens, Detecting atrange attractors in turbulence, in: D. Rand, L. Young (Eds.), Dynamical Systems and Turbulence, Warwick, in: Lecture Notes in Mathematics, vol. 898, Springer-Verlag, Berlin, 1981, pp. 366–381.
[66] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time-series, Physica D 16 (1984) 285–317.