Nonlinearity **19** (2006) 1801–1847 doi:10.1088/0951-7715/19/8/005

Structural stability and hyperbolicity violation in high-dimensional dynamical systems

D J Albers[1,2,3,4] **and J C Sprott**[2]

[1] Max Planck Institute for Mathematics in the Science, Leipzig 04103, Germany
[2] Physics Department, University of Wisconsin, Madison, WI 53706, USA
[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[4] Computational Science and Engineering Center, University of California, Davis, One Shields Ave, Davis, CA 95616, USA

E-mail: albers@cse.ucdavis.edu and sprott@physics.wisc.edu

Received 24 October 2005, in final form 6 June 2006
Published 7 July 2006
Online at stacks.iop.org/Non/19/1801

Recommended by B Eckhardt

Abstract

This report investigates the dynamical stability conjectures of Palis and Smale and Pugh and Shub from the standpoint of numerical observation and lays the foundation for a stability conjecture. As the dimension of a dissipative dynamical system is increased, it is observed that the number of positive Lyapunov exponents increases monotonically, the Lyapunov exponents tend towards continuous change with respect to parameter variation, the number of observable periodic windows decreases (at least below numerical precision) and a subset of parameter space exists such that topological change is very common with small parameter perturbation. However, this seemingly inevitable topological variation is never catastrophic (the dynamic type is preserved) if the dimension of the system is high enough.

PACS numbers: 05.45.−a, 89.75.−k, 05.45.Tp, 89.70.+c, 89.20.Ff

1. Introduction

Much of the work in the fields of dynamical systems and differential equations has, for the last hundred years, entailed the classification and understanding of the qualitative features of the space of differentiable mappings. A primary focus is the classification of topological differences between different systems (e.g. structural stability theory). One of the primary difficulties of such a study is choosing a notion of behaviour that is not so strict such that it differentiates on too trivial a level, yet is strict enough that it has some meaning (e.g. the Palis–Smale stability conjecture uses topological equivalence whereas the Pugh–Shub stability conjecture uses stable ergodicity). Most stability conjectures are with respect to any C^r (r varies from conjecture to conjecture) perturbation in the set of bounded C^k diffeomorphisms which allows for variation of the functional form of the mapping with respect to the Whitney C^r topology (there are no real notions of parameters in a practical sense). In this framework, perturbations refer to perturbations of the graph of the mapping—from a practical standpoint infinitely many 'parameters' are needed for arbitrary perturbations of the graph of the mapping. This tactic is employed both for ease of generalization to many mappings and because of the geometrical argument styles that characterize mathematical dynamical systems. This differs from the more practical construction used by the non-mathematics dynamics community where nearly all dynamical systems studied have a finite number of parameters. We will concern ourselves with a practical construction where we vary a finite number of parameters—yet study mappings that are 'universal approximators' and can approximate to arbitrary accuracy general C^r mappings and their derivatives in the limit where there exist infinitely many parameters. Unlike much work involving stability conjectures, our work is *numerical* and it focuses on *observable* asymptotic behaviour in high-dimensional systems. Our chief claim is that generally, for high-dimensional dynamical systems in our construction, there exist large portions of parameter space such that topological variation inevitably accompanies parameter variation, yet the topological variation happens in a 'smooth', non-erratic manner. Let us state our results without rigour or particular precision, noting that we will save more precise statements for section 3.

Statement of results 1 (Informal). *Given our particular impositions (sections 2.1.4 and 2.1.1) upon a space of C^r discrete-time maps from compact sets to themselves relative to a measure and an invariant (SRB) measure (used for calculating Lyapunov exponents), in the limit of high dimension, there exists a subset of parameter space such that strict hyperbolicity is violated (implying changes in the topological structure) on a nearly dense (and hence unavoidable), yet low-measure (with respect to Lebesgue measure), subset of parameter space.*

A more refined version of this statement will contain all of our results. For mathematicians, we note that although the stability conjecture of Palis and Smale [1, 2] is quite true (as proved by Robbin [3], Robinson [4] and Mañé [5]), we show that in high dimensions this structural stability may occur over such small sets in the parameter space that it may never be observed in chaotic regimes of parameter space. Nevertheless, this lack of observable structural stability has very mild consequences for applied scientists.

1.1. Outline

As this paper is attempting to reach a diverse readership, we will briefly outline the work for ease of reading. Of the remaining introduction sections, section 1.2 can be skipped by readers familiar with the stability conjecture of Smale and Palis, the stable ergodicity of Pugh and Shub and the results from previous computational studies.

Following the introduction we will address various preliminary topics pertaining to this report. Beginning in section 2.1.1, we present mathematical justification for using time-delay maps for a general study of $d > 1$ dimensional dynamical systems. This section is followed by a discussion of neural networks, beginning with their definition in the abstract (section 2.1.2). Following the definition of neural networks, we carefully specify the mappings that neural networks are able to approximate (section 2.1.3), and by doing so, the space of mappings we study. In section 2.1.4 we give the specific construction (with respect to the neural networks) we will use in this paper. In section 2.1.5 we tie sections 2.1.1–2.1.4 together. Those uninterested in the mathematical justifications for our models and only interested in our specific formulation should skip sections 2.1.1 through 2.1.3 and concentrate on sections 2.1.4 and 2.1.5. The discussion of the set of mappings we will study is followed by relevant definitions from hyperbolicity and ergodic theory (section 2.2). It is here where we define the Lyapunov spectrum, hyperbolic maps and discuss relevant stability conjectures. Section 2.3 provides justification for our use of Lyapunov exponent calculations on our space of mappings (the neural networks). Readers familiar with topics in hyperbolicity and ergodic theory can skip this section and refer to it as is needed for an understanding of the results. Lastly, in section 2.4, we make a series of definitions we will need for our numerical arguments. Without an understanding of these definitions, it is difficult to understand both our conjectures and our arguments.

Section 3 discusses the conjectures we wish to investigate formally. For those interested in just the results of this report, reading sections 2.4, 3 and 7 will suffice. The next section, section 4, discusses the errors present in our chief numerical tool, the Lyapunov spectrum. This section is necessary for a fine and careful understanding of this report, but this section is easily skipped upon first reading. We then begin our preliminary numerical arguments. Section 5 addresses the three major properties we need to argue for our conjectures. For an understanding of our arguments and why our conclusions make sense, reading this section is necessary. The main arguments regarding our conjectures follow in section 6. It is in this section that we make the case for the claims of section 3. The summary section (section 7) begins with a summary of our numerical arguments and how they apply to our conjectures. We then interpret our results in light of various stability conjectures and other results from the dynamics community.

1.2. Background

The goal of this paper is three-fold. First, we want to construct a framework for pursuing a qualitative, numerical study of dynamical systems. Second, utilizing diagnostics available to numericists, we want to analyse and characterize behaviour of the space of functions constrained by a measure on the respective function space. Finally, we want to relate, to the extent possible, the qualitative features we observe with results from the qualitative dynamical systems theory. To set the stage for this, and because the mathematical dynamics results provide the primary framework, language set and motivation for this study, it is important to understand both the development of the qualitative theory of dynamical systems as constructed by mathematicians as well as numerical results for constructions similar to the one we will utilize.

From the perspective of mathematics, the qualitative study of dynamical systems was devised by Poincaré [6] and was motivated primarily by a desire to classify dynamical systems by their topological behaviour. The hope was that an understanding of the space of mappings used to model nature would provide insight into nature (see [7–11]). For physicists and other scientists, a qualitative understanding is three-fold. First, the mathematical dynamics construction provides a unifying, over-arching framework—the sort of structure necessary for building and interpreting dynamics theory. Second, the analysis is rooted in manipulation of graphs and, in general, the geometry and topology of the dynamical system which is independent of, and yet can be related to, the particular scientific application. It is useful to have a geometric understanding of the dynamics. Third, from a more practical perspective, most experimentalists who work on highly nonlinear systems (e.g. plasma physics and fluid dynamics) are painfully aware of the dynamic stability that mathematicians eventually hope to capture with stability conjectures or theorems. Experimentalists have been attempting to control and eliminate complex dynamical behaviour since they began performing experiments—it is clear from experience that dynamic types such as turbulence and chaos are highly stable with respect to perturbations in highly complicated dynamical systems; the why and how of the dynamic stability, and defining the right notion of equivalence to capture that stability, is the primary difficult question. The hope is that, if the geometric characteristics that allow chaos to persist can be understood, it might be easier to control or even eliminate those characteristics. At the very least, it would be useful to know very precisely why we cannot control or rid our systems of complex behaviour.

In the 1960s, to formulate and attack a stability conjecture (see [12,13]) that would achieve a qualitative description of the space of dynamical systems, Smale introduced axiom A(nosov). Dynamical systems that satisfy axiom A are strictly hyperbolic (definitions (6) and (7)) and have dense periodic points on the non-wandering set[5]. A further condition that was needed was the strong transversality condition—f satisfies the strong transversality condition when, for every $x \in M$, the stable and unstable manifolds W_x^s and W_x^u are transverse at x. Using axiom A and the strong transversality condition, Palis and Smale stated what they hoped would be a relatively simple qualitative characterization of the space of dynamical systems using structural stability (i.e. topological equivalence (C^0) under perturbation, see 'volume 14' [1], which contains many crucial results). The conjecture, stated most simply, says,'a system is C^r stable if its limit set is hyperbolic and, moreover, stable and unstable manifolds meet transversally at all points' [9]. That axiom A and strong transversality imply C^r structural stability was shown by Robbin [3] for $r \geqslant 2$ and Robinson [4] for $r = 1$. The other direction of the stability conjecture was much more elusive, yet in 1980 this was shown by Mañé [5] for $r = 1$. Nevertheless, due to many examples of structurally unstable systems being dense amongst many 'common' types of dynamical systems, proposing some global structure for a space of dynamical systems became much more unlikely (see 'volume 14' [1] for several counter-examples to the claims that structurally stable dynamical systems are dense in the space of dynamical systems). Newhouse [14] was able to show that infinitely many sinks occur for a residual subset of an open set of C^2 diffeomorphisms near a system exhibiting a homoclinic tangency. Further, it was discovered that orbits can be highly sensitive to initial conditions [15–18]. Much of the sensitivity to initial conditions was investigated numerically by non-mathematicians. Together, the examples from both pure mathematics and the sciences sealed the demise of a simple characterization of dynamical systems by their purely topological properties. Nevertheless, despite the fact that structural stability does not capture all that we wish, it is still a very useful, intuitive tool.

Again, from a physical perspective, the question of the existence of dynamic stability is not open—physicists and engineers have been trying to suppress chaos and turbulence in high-dimensional systems for several hundred years. In fact, simply being able to make a consistent measurement implies persistence of a dynamic phenomenon. The trick in mathematics is writing down a relevant notion of dynamic stability and then the relevant necessary geometrical characteristics to guarantee dynamic stability. From the perspective of modelling nature, the aim of structural stability was to imply that if one selects (fits) a model equation, small errors will be irrelevant since small C^r perturbations will yield topologically equivalent models. This did not work because topological equivalence is too strong a specification of equivalence for structural stability to apply to the broad range of systems we wish it to apply to. In particular, it was strict hyperbolicity that was the downfall of the Palis–Smale stability conjecture because non-hyperbolic dynamical systems could be shown to persist under perturbations. To combat this, in the 1990s Pugh and Shub [19] introduced the notion of stable ergodicity and with it a stability conjecture that includes measure-theoretic properties required for dynamic stability

[5] $\Omega(f) = \{x \in M | \forall$ neighbourhood U of $x, \exists n \geqslant 0$ such that $f^n(U) \cap U \neq 0\}$.

and a weakened notion of hyperbolicity introduced by Brin and Pesin [20]. A thesis for their work was [19], 'a little hyperbolicity goes a long way in guaranteeing stably ergodic behaviour'. This thesis has driven the partial hyperbolicity branch of dynamical systems and is our claim as well. A major thrust of this paper is to, in a practical, computational context, investigate the extent to which ergodic behaviour and topological variation (versus parameter variation) behave given a 'little bit' of hyperbolicity. Further, we will formulate a method to study, and subsequently *numerically* investigate, one of the overall haunting questions in dynamics: how much of the space of bounded C^r ($r > 0$) systems is hyperbolic, and how many of the pathologies found by Newhouse and others are observable (or even existent) in the space of bounded C^r dynamical systems, all relative to a measure. The key to this proposal is a computational framework utilizing a function space that can approximate C^r and admits a measure relative to which the various properties can be specified.

One of the early numerical experiments of a function space akin to the space we study was performed by Sompolinksky *et al* [21], who analysed 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 such that 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. Often, these findings are argued using random matrix theory using results of Girko *et al* [22], Edelman [23] and Bai [24]. In a similar vein, Doyon *et al* [25, 26] 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 also justified their findings with results from random matrix theory. Cessac *et al* [27] came to similar conclusions with slightly different networks and 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. A primary conclusion of Cessac [28], a conclusion that was proved in [29] by an analysis given the local chaos hypothesis, was that in the space of neural networks in the Cessac construction, the degrees of freedom become independent, and thus the dynamics resemble Brownian motion.

The network architecture utilized in this paper is fundamentally different from the architectures used by Sompolinsky, Cessac or Doyon. In particular, the aforementioned studies use networks that map a compact set in R^d to itself, and there is no time-delay. This paper utilizes feed-forward, discrete time, *time-delayed* neural networks. The difference in the respective architectures induces some very important differences. The local chaos hypothesis, the property that allows the mean field analysis in the aforementioned studies, is not valid for time-delayed networks. Moreover, the construction we utilize induces correlations in time and space, albeit in a uniform, random manner; the full effects of which are a topic of ongoing work. However, current intuition suggests that the chaotic regime of the systems we study have turbulent-like dynamics rather than Brownian-motion-like dynamics. The choice of time-delay neural networks was motivated by the relative ease of computation for the derivative map, the approximation theorems of Hornik *et al* [30], the fact that time-delay neural networks are a very practical tool for time-series reconstruction and that the embedology (and the associated notion of prevalence) theory of Sauer *et al* [31] can be used to construct a means of connecting abstract and computational results. It is likely that all these systems can be shown to be equivalent via an embedding, but the implied measures on the respective spaces of mappings could be (and probably are) rather different. In the end, it is reassuring that many of the results for the different architectures are often similar despite the networks being fundamentally different.

Other noteworthy computational studies of similar flavour, but using other frameworks, can be found in [18,32,33].

Finally, general background regarding the dynamics of the particular space of mappings we utilize can be found in [34] and [35]. In particular, studies regarding bifurcations from fixed points and the route to chaos can be found in [36–38], respectively. In this work we will discuss, vaguely, taking high-dimension, high-number-of-parameter limits. This can be a very delicate matter, and some of the effects of increasing the number of dimensions and the number of parameters are addressed in [34].

2. Definitions and preliminaries

In this section we will define the following items: the family of dynamical systems we wish to investigate, the function space we will use in our experiments, Lyapunov exponents and definitions specific to our numerical arguments.

2.1. The space of mappings

The motivation and construction of the set of mappings we will use for our investigation of dynamical systems follows from the embedology of Sauer *et al* [31] (see [39,40] for the Takens construction) and the neural network approximation theorems of Hornik *et al* [30]. We will use embedology to demonstrate how studying time-delayed maps of the form $f : R^d \to R$ is a natural choice for studying standard dynamical systems of the form $F : R^d \to R^d$. In particular, embedding theory shows an equivalence via the approximation capabilities of scalar time-delay dynamics with standard, $x_{t+1} = F(x_t)$ ($x_i \in R^d$) dynamics. However, there is no mention of, in a practical sense, the explicit functions used for the reconstruction of the time-series in the Takens or the embedology construction. The neural network approximation results show in a precise and practical way what a neural network is and what functions it can approximate ([41]). In particular, it says that neural networks can approximate the $C^r(R^d)$ mappings and their derivatives (and indeed are dense in $C^r(R^d)$ on compacta), but there is no mention of the time-delays or reconstruction capabilities we use. Thus, we need to discuss both the embedding theory and the neural network approximation theorems.

Those not interested in the justification of our construction may skip to section 2.1.4 where we define, in a concrete manner, the neural networks we will employ.

2.1.1. Dynamical systems construction. In this paper we wish to investigate dynamical systems mapping compact sets (attractors) to themselves. Specifically, begin with an open set $U \subset R^q$ ($q \in N$), a compact invariant set A such that $A \subset U \subset R^q$ where $\mathrm{boxdim}(A) = d \leqslant q$ and a diffeomorphism $F \in C^r(U)$ for $r \geqslant 2$ defined as

$$x_{t+1} = F(x_t) \tag{1}$$

with $x_t \in U$. However, for computational reasons, and because practical scientists work with scalar time-series data, we will be investigating this space with neural networks that can approximate (see section 2.1.3) dynamical systems $f \in C^r(R^d, R)$ which are time-delay maps given by

$$y_{t+1} = f(y_t, y_{t-1}, \ldots, y_{t-(d-1)}), \tag{2}$$

where $y_t \in R$. Both systems (1) and (2) form dynamical systems. However, since we intend to use systems of the form (2) to investigate the space of dynamical systems as given in equation (1), we must show how mappings of the form (2) are related to mappings of the form (1).

The relationship between a dynamical system and its delay dynamical system is not one of strict equivalence; the relationship is via an embedding. In general, we call $g \in C^k(U, R^n)$ an embedding if $k \geqslant 1$ and if the local derivative map (the Jacobian—the first order term in the Taylor expansion) is one-to-one for every point $x \in U$ (i.e. g must be an immersion). The idea of an embedding theorem is that, given a d-dimensional dynamical system and a 'measurement function', $E : A \to R$ (E is a C^k map), where E represents some empirical style measurement of F, there is a 'Takens map' (which does the embedding) g for which $x \in A$ can be represented as a $2d+1$ tuple $(E(x), E \circ F(x), E \circ F^2(x), \ldots, E \circ F^{2d}(x))$ where F is an ordinary difference equation (time evolution operator) on A. Note that the $2d + 1$ tuple is a time-delay map of x. The map that iterates the time-delay coordinates, denoted the *delay coordinate map*, is given by $\tilde{F} = g \circ F \circ g^{-1}$. There is no strict equivalence between dynamical systems and their time-delay maps because many different measurement functions exist that yield an embedding of A in R^{2d+1}, thus destroying any one-to-one relationship between equations (1) and (2). Whether typical measurement functions will yield an embedding (and thus preserve the differential structure) was first undertaken by Takens [39, 40]. In the Takens construction, the original mapping (F) is of C^k manifolds to themselves, not attractors with fractal dimensions, and the notion of typical is of (topological) genericity (i.e. a countable union of open and dense sets) that is compatible with C^r where no meaningful notion of measure can be directly imposed. This framework was generalized by Sauer *et al* [31, 42] to a more practical setting by devising a practical, measure-theoretic notion of typical, and proving results applicable to invariant sets (attractors) instead of just manifolds. The key to the generalization of the notion of typical from generic to a measure-theoretic notion lies in the construction of the embedding. An embedding, in the particular circumstance of interest here, couples a space of dynamical systems with a space of measurement functions. In the Takens construction, the embedding was $g \in S \subset \mathrm{Diff}(M) \times C^k(M, R)$ where $F \in \mathrm{Diff}(M)$ and the measurement function was in $E \in \mathcal{E} \subset C^k(M, R)$. Because $\mathcal{E} \subset C^k(M, R)$, the only notion available to specify typical is genericity. However, the measurement function space, $\mathcal{E}$, can be replaced with a (finite-dimensional) function space that can yield a measure—such as polynomials of degree up to $2d + 1$ with d variables. Equipped with a measurement function space that yields a measure, a new notion of typical, devised in [31] and discussed in [43–45], *prevalence*, can be defined and utilized.

Definition 1 (Prevalence [31]). *A Borel subset S of a normed linear space V is prevalent if there is a finite-dimensional subspace $\mathcal{E} \subset V$ such that for each $v \in V$, $v + e \in S$ for Lebesgue a.e. $e \in \mathcal{E}$.*

Note that $\mathcal{E}$ is referred to as the probe space. In the present setting, V represents F and the associated space of measurement functions, E, is $\mathcal{E}$. Thus, $\mathcal{E}$ is a particular finite-dimensional space of (practical) measurement functions. The point is that S is prevalent if, for any starting point in V, variation in E restricted to $\mathcal{E}$ will remain in S Lebesgue a.e. We can now state the embedding theorem relevant to the construction we are employing.

Theorem 1. *Let $F \in C^r(U)$, $r \geqslant 1$, $U \subset R^q$, and let A be a compact subset of U with boxdim$(A) = d$, and let $w > 2d$, $w \in N$ (often $w = 2d + 1$). Assume that for every positive integer $p \leqslant w$, the set A_p of periodic points of period p satisfied boxdim$(A_p) < p/2$, and that the linearization DF^p for each of these orbits has distinct eigenvalues. Then, for almost every (in the sense of prevalence) C^1 measurement function E on U, $\tilde{F} : U \to R^w$ is (i) one-to-one on A, and (ii) is an immersion on the compact subset C of a smooth manifold contained in A.*

This thus defines an equivalence, that of an embedding (the Takens map, $g : U \to R^w$), between time-delayed Takens maps of 'measurements' and the 'actual' dynamical system

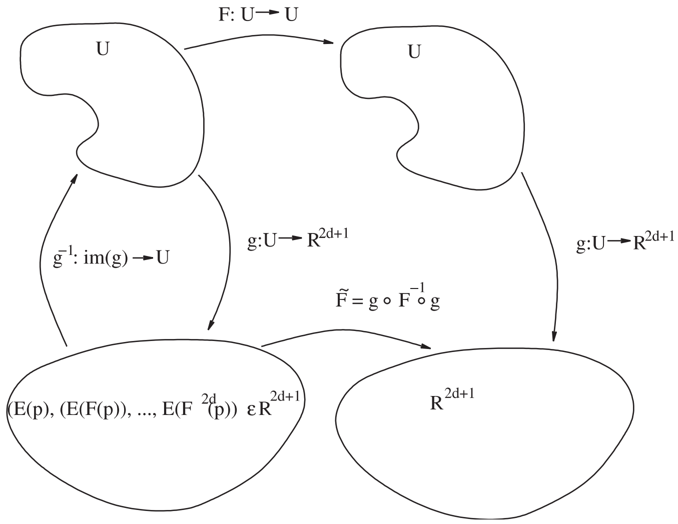

Figure 1. Schematic diagram of the embedding theorems applied to our construction.

operating in time on $x_t \in U$. This does not imply that all time-delay maps can be embeddings of a C^r dynamical system (see [31] for a detailed discussion of how time-delay dynamical systems can fail to yield an embedding).

To explicitly demonstrate how this applies to our circumstances, consider figure 1 in which F and E are as given above and the embedding g is explicitly given by

$$g(x_t) = (E(x_t), E(F(x_t)), \ldots, E(F^{2d}(x_t))). \tag{3}$$

In a colloquial, experimental sense, $\tilde{F}$ just keeps track of the observations from the measurement function E and, at each time step, shifts the newest observation into the $w = 2d+1$ tuple and sequentially shifts the scalar observation at time t (y_t) of the $2d + 1$ tuple to the $t - 1$ position of the $2d + 1$ tuple. In more explicit notation, $\tilde{F}$ is the following mapping:

$$(y_1, \ldots, y_{2d+1}) \mapsto (y_2, \ldots, y_{2d+1}, g(F(g^{-1}(y_1, \ldots, y_{2d+1})))), \tag{4}$$

where, again, $\tilde{F} = g \circ F \circ g^{-1}$. The embedology theorem of Sauer *et al* says that there is a prevalent Borel subset (in a probe space, S) of measurement functions (E) of dynamical systems of the form (1) (F), which will yield an embedding g of A. The neural networks we will propose in the sections that follow can approximate $\tilde{F}$ and its derivatives (to any order) to arbitrary accuracy (a notion we will make more precise later). Thus, we will investigate the space of C^r dynamical systems given by (1) via the mappings used to approximate them.

2.1.2. Abstract neural networks. Begin by noting that, in general, a feed-forward, scalar neural network is a C^r mapping $\gamma : R^d \to R$; the set of feed-forward, scalar networks with a single hidden layer, $\Sigma(G)$, can be written as:

$$\Sigma(G) \equiv \{\gamma : R^d \to R | \gamma(x) = \sum_{i=1}^{N} \beta_i G(\tilde{x}^T \omega_i)\}, \tag{5}$$

where $x \in R^d$ is the d-vector of network inputs, $\tilde{x}^T \equiv (1, x^T)$ (where x^T is the transpose of x), N is the number of hidden units (neurons), $\beta_1, \ldots, \beta_N \in R$ are the hidden-to-output layer weights, $\omega_1, \ldots, \omega_N \in R^{d+1}$ are the input-to-hidden layer weights and $G : R^d \to R$ is the hidden layer activation function (or neuron). The partial derivatives of the network output function, γ, are

$$\frac{\partial \gamma(x)}{\partial x_k} = \sum_{i=1}^{N} \beta_i \omega_{ik} DG(\tilde{x}^T \omega_i), \tag{6}$$

where x_k is the kth component of the x vector, ω_{ik} is the kth component of ω_i and DG is the usual first derivative of G. The matrix of partial derivatives (the Jacobian) takes a particularly simple form when the x vector is a sequence of time delays ($x_t = (y_t, y_{t-1}, \ldots, y_{t-(d-1)})$ for $x_t \in R^d$ and $y_i \in R$). It is precisely for this reason that the time-delayed formulation eases the computational woes.

This paper contains numerical evidence of a geometric mechanism, specified by the Lyapunov spectrum, that is present in the above class of functions as $d \to \infty$. In doing so, we will often make arguments that involve taking $d \to \infty$ while leaving N limits vague. That we can do this is a result of the function approximation characteristics of the neural networks; to see N specified in particular see [34, 46].

2.1.3. Neural networks as function approximations. Hornik *et al* [30] provided the theoretical justification for the use of neural networks as function approximators. The aforementioned authors provide a degree of generality that we will not need; for their results in full generality see [30, 47].

The ability of neural networks to approximate (and be dense in) the space of mappings relevant to dynamics is the subject of the keynote theorems of Hornik *et al* [30] (for their results in full generality see [30, 47]). However, to state these theorems, a discussion of Sobolev function space, S_p^m, is required. We will be brief, noting references Adams and Fourier [48] and Hebey [49] for readers wanting more depth with respect to Sobolev spaces. For the sake of clarity and simplification, let us make a few remarks which will pertain to the rest of this section:

(i) μ is a measure; λ is the standard Lebesgue measure; for all practical purposes, $\mu = \lambda$;

(ii) l, m and d are finite, non-negative integers; m will be with reference to a degree of continuity of some function spaces and d will be the dimension of the space we are operating on;

(iii) $p \in R$, $1 \leqslant p < \infty$; p will be with reference to a norm—either the L_p norm or the Sobolev norm;

(iv) $U \subset R^d$, U is measurable.

(v) $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)^T$ is a d-tuple of non-negative integers (or a multi-index) satisfying $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, $|\alpha| \leqslant m$;

(vi) for $x \in R^d$, $x^\alpha \equiv x_1^{\alpha_1} \cdot x_2^{\alpha_2}, \ldots, x_d^{\alpha_d}$;

(vii) D^α denotes the partial derivative of order $|\alpha|$

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha} \equiv \frac{\partial^{|\alpha|}}{(\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_d^{\alpha_d})}; \tag{7}$$

(viii) $u \in L^1_{loc}(U)$ is a locally integrable, real valued function on U;

(ix) $\rho^m_{p,\mu}$ is a metric, dependent on the subset U, the measure μ, and p and m in a manner we will define shortly;

(x) $\| \cdot \|_p$ is the standard norm in $L_p(U)$;

Letting m be a positive integer and $1 \leqslant p < \infty$, we define the Sobolev norm, $\| \cdot \|_{m,p}$, as follows:

$$||u||_{m,p} = \left(\sum_{0 \leqslant |\alpha| \leqslant m} (\| D^{\alpha} u \|_p^p) \right)^{1/p}, \tag{8}$$

where $u \in L^1_{loc}(U)$ is a locally integrable, real valued function on $U \subset R^d$ (u could be significantly more general) and $|| \cdot ||_p$ is the standard norm in $L_p(U)$. Likewise, the Sobolev metric can be defined:

$$\rho^m_{p,\mu}(f, g) \equiv ||f - g||_{m,p,U,\mu}. \tag{9}$$

It is important to note that this metric is dependent on U.

For ease of notation, let us define the set of m-times differentiable functions on U,

$$C^m(U) = \{f \in C(U) | D^{\alpha} f \in C(U), ||D^{\alpha} f||_p < \infty \; \forall \alpha, |\alpha| \leqslant m\}. \tag{10}$$

We are now free to define the Sobolev space for which our results will apply.

Definition 2. *For any positive integer m and $1 \leqslant p < \infty$, we define a Sobolev space $S^m_p(U, \lambda)$ as the vector space on which $|| \cdot ||_{m,p}$ is a norm:*

$$S^m_p(U, \lambda) = \{f \in C^m(U) | \; ||D^{\alpha} f||_{p,U,\lambda} < \infty \quad \textit{for all } |\alpha| \leqslant m\} \tag{11}$$

Equipped with the Sobolev norm, S^m_p is a Sobolev space over $U \subset R^d$.

Two functions in $S^m_p(U, \lambda)$ are close in the Sobolev metric if all the derivatives of order $0 \leqslant |\alpha| < m$ are close in the L_p metric. It is useful to recall that we are attempting to approximate $\tilde{F} = g \circ F \circ g^{-1}$ where $\tilde{F} : R^{2d+1} \to R$; for this task the functions from $S^m_p(U, \lambda)$ will serve quite nicely. The whole point of all this machinery is to state approximation theorems that require specific notions of density. Otherwise we would refrain and instead use the standard notion of C^k functions—the functions that are k-times differentiable uninhibited by a notion of a metric or norm.

Armed with a specific function space for which the approximation results apply (there are many more), we will conclude this section by briefly stating one of the approximation results. However, before stating the approximation theorem, we need two definitions—one which makes the notion of closeness of derivatives more precise and one which gives the sufficient conditions for the activation functions to perform the approximations.

Definition 3 (m-uniformly dense). *Assume m and l are non-negative integers $0 \leqslant m \leqslant l$, $U \subset R^d$ and $S \subset C^l(U)$. If for any $f \in S$, compact $K \subset U$ and $\epsilon > 0$ there exists a $g \in \Sigma(G)$ such that*

$$\max_{|\alpha| \leqslant m} \sup_{x \in K} |D^{\alpha} f(x) - D^{\alpha} g(x)| < \epsilon \tag{12}$$

then $\Sigma(G)$ is m-uniformly dense on compacta in S.

It is this notion of m-uniformly dense in S that provides all the approximation power of both the mappings and the derivatives (up to order l) of the mappings. Next we will supply the condition on our activation function necessary for the approximation results.

Definition 4 (l-finite). *Let l be a non-negative integer. G is said to be l-finite for $G \in C^l(R)$ if*

$$0 < \int |D^l G| \mathrm{d}\lambda < \infty, \tag{13}$$

i.e. the lth derivative of G must be both bounded away from zero, and finite for all l (recall $\mathrm{d}\lambda$ *is the standard Lebesgue volume element).*

The hyperbolic tangent, our activation function, is l-finite.

With these two notions, we can state one of the many existing approximation results.

Corollary 1. (corollary 3.5 [30]) *If G is l-finite, $0 \leqslant m \leqslant l$, and U is an open subset of R^d, then $\Sigma(G)$ is m-uniformly dense on compacta in $S_p^m(U, \lambda)$ for $1 \leqslant p < \infty$.*

In general, we wish to investigate differentiable mappings of compact sets to themselves. Further, we wish for the derivatives to be finite almost everywhere. Thus, the space $S_p^m(U, \lambda)$ will suffice for our purposes. Our results also apply to piecewise differentiable mappings. However, this requires a more general Sobolev space, W_p^m. We have refrained from delving into the definition of this space since it requires a bit more formalism. For those interested see [30] and [48].

2.1.4. Explicit neural network construction. The single layer feed-forward neural networks (γ from the above section) we will consider are of the form

$$x_t = \beta_0 + \sum_{i=1}^{N} \beta_i G \left(s\omega_{i0} + s \sum_{j=1}^{d} \omega_{ij} x_{t-j} \right), \tag{14}$$

which is a map from R^d to R. The squashing function G, for our purpose, will be the hyperbolic tangent. In (14), N represents the number of hidden units or neurons, d is the input or embedding dimension of the system which functions simply as the number of time lags and s is a scaling factor on the weights.

The parameters are real ($\beta_i, \omega_{ij}, x_j, s \in R$) and the β_i and ω_{ij} are elements of weight matrices (which we hold fixed for each case). The initial conditions are denoted $(x_0, x_1, \ldots, x_d)$, and $(x_t, x_{t+1}, \ldots, x_{t+d})$ represents the current state of the system at time t. The βs are iid uniform over [0, 1] and then re-scaled to satisfy the condition $\sum_{i=1}^{N} \beta_i^2 = N$. The ω_{ij}s are iid normal with zero mean and unit variance. 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 are chosen iid uniform on the interval $[-1, 1]$. All the weights and initial conditions are selected randomly using a pseudo-random number generator [50, 51].

We would like to make a few notes with respect to the squashing function, tanh(). First, $\tanh(x)$, for $|x| \gg 1$, will tend to behave much like a binary function. Thus, the states of the neural network will tend towards the finite set $(\beta_0 \pm \beta_1 \pm \beta_2 \ldots \pm \beta_N)$, or a set of 2^N different states. In the limit where the arguments of tanh() become infinite, the neural network will have periodic dynamics. Thus, if $\langle\beta\rangle$ or s become very large, the system will have a greatly reduced dynamic variability. Based on this problem, one might feel tempted to bound the β *a la* $\sum_{i=1}^{N} |\beta_i| = k$ fixing k for all N and d. This is a bad idea, however, since, if the β_i are restricted to a sphere of radius k, as N is increased, $\langle{\beta_i}^2\rangle$ goes to zero [52]. The other extreme of tanh() also yields a very specific behaviour type. For x very near 0, the $\tanh(x)$ function is nearly linear. Thus, choosing s small will force the dynamics to be mostly linear, again yielding fixed point and periodic behaviour (no chaos). Thus the scaling parameter s provides a unique bifurcation parameter that will sweep from linear ranges to highly non-linear ranges, to binary ranges—fixed points to chaos and back to periodic phenomena.

2.1.5. Neural network approximation theorems and the implicit measure. Because the neural networks can be identified with points in $R^k (k = N(d+2)+2)$, we impose a probability measure on $\Sigma(\tanh)$ by imposing a probability measure on parameter space R^k. These measures, which were explicitly defined in section 2.1.4, can be denoted m_β (on the β), m_ω (on the ω), m_s (on s), and m_I (on the initial conditions) and form a product measure on $R^k \times R^d$ in the standard way. All the results in this paper are subject to this measure—this is part of the point of this construction. We are not studying a particular dynamical system, but rather we are performing a statistical study of a *function space*. From the measure-theoretic perspective, all networks are represented with this measure, although clearly not with equal likelihood. For instance, Hamiltonian systems are likely a zero measure set relative to the measures we utilize. Moreover, the joint-probability distributions characteristic of neural networks, where training has introduced correlations between weights, will likely produce greatly varying results depending on the source of the training (for analytical tools for the analysis of trained networks, see [53]).

Specifying 'typical', or forging a connection between the results we present and the results of other computational studies, or the results from abstract dynamics is a delicate problem. Most abstract dynamics results are specified for the C^r function space in the C^k Whitney topology using the notion of a *residual* (generic) subset because no sensible measure can be imposed on the infinite-dimensional space of C^r systems. Because we are utilizing a space that, with finitely many parameters, is a finite-dimensional approximation of C^r, the space we study will always be non-residual relative to C^r. In fact, even when m-uniformal density is achieved at $N = \infty$, C^r still has an uncountably infinite number of dimensions as opposed to the countably infinite number of dimensions induced by m-uniformal density on $\Sigma(\tanh)$. Moreover, because we are studying a space of functions and not a particular mapping, comparing results of, say, coupled-map lattices with our study must be handled with great care. A full specification of such relationships, which we will only hint at here, will involve two key aspects. First, the proper language to relate the finite-dimensional neural network construction to the space of embeddings of dynamical systems is that of prevalence [43–45]. Prevalence (see definition 1) provides a measure-theoretic means of coping with the notion of typical between finite and infinite-dimensional spaces. Second, training the neural networks induces (joint) probability measures on the weight structure and thus a means of comparison in measure of various different qualitative dynamics types (see [53] as a starting point). In the simplest situation, if the measures of two trained ensembles of networks are singular with respect to each other, it is likely that the sets containing the volume as determined by the measures are disjoint. Thus, the space, $\Sigma(\tanh)$, with different measures can be used to quantify, in a precise way, the relationship between different dynamical types.

In a practical sense, this construction is easily extendible in various ways. One extension of particular relevance involves imposing different measures on each component of R^k producing a more non-uniform connection structure between the states (a different network structure). Another extension involves imposing a measure via training [41, 54, 55] on, for instance, coupled-map data (or data from any other common mapping). Because the latter situation induces a dependence between the weights, thus R^k will be equipped with a joint probability measure rather than a simple product measure. Given the neural network approximation theorems, imposing different measures, and in particular, correlations between the weights, should have the same effect that studying different physical systems will have. The point is, with the abstract framework outlined above, there is a way of classifying and relating systems by their dynamics via the space of measures on the parameter space of neural networks.

2.2. *Characteristic Lyapunov exponents, hyperbolicity, and structural stability*

The theories of hyperbolicity, Lyapunov exponents and structural stability have had a long, wonderful and tangled history beginning with Smale [56] (for other good starting points see [9, 11, 13, 57]). We will, of course, only scratch the surface with our current discussion, but rather put forth the connections relevant for our work. In structural stability theory, topological equivalence is the notion of equivalence between dynamical systems, and structural stability is the notion of the persistence of topological structure.

Definition 5 (Structural stability). *A C^r discrete-time map, f, is structurally stable if there is a C^r neighbourhood, V of f, such that any $g \in V$ is topologically conjugate to f, i.e. for every $g \in V$, there exists a homeomorphism h such that $f = h^{-1} \circ g \circ h$.*

In other words, a map is structurally stable if, for all other maps g in a C^r neighbourhood, there exists a homeomorphism that will map the domain of f to the domain of g, the range of f to the range of g and the inverses, respectively. Structural stability is a purely topological notion. The hope of Smale and others was that structural stability would be open-dense (generic) in the space of C^r dynamical systems as previously noted in section 1.2 (see [1] for counter-examples). Palis and Smale then devised the stability conjecture that would fundamentally tie together hyperbolicity and structural stability.

The most simple and intuitive definition of hyperbolic applies to the linear case.

Definition 6 (Hyperbolic linear map). *A linear map of R^n is called hyperbolic if all of its eigenvalues have modulus different from one.*

Because strict hyperbolicity is a bit restrictive, we will utilize the notion of uniform partial hyperbolicity which will make precise the notion of a 'little bit' of hyperbolicity.

Definition 7 (Partial hyperbolicity). *The diffeomorphism f of a smooth Riemannian manifold M is said to be partially hyperbolic if for all $x \in M$ the tangent bundle T_xM has the invariant splitting:*

$$T_xM = E^u(x) \oplus E^c(x) \oplus E^s(x) \tag{15}$$

into strong stable $E^s(x) = E^s_f(x)$, strong unstable $E^u(x) = E^u_f(x)$ and central $E^c(x) = E^c_f(x)$ bundles, at least two of which are non-trivial[6]. Thus, there will exist numbers $0 < a < b < 1 < c < d$ such that, for all $x \in M$:

$$v \in E^u(x) \Rightarrow d||v|| \leqslant ||D_x f(v)||, \tag{16}$$

$$v \in E^c(x) \Rightarrow b||v|| \leqslant ||D_x f(v)|| \leqslant c||v||, \tag{17}$$

$$v \in E^s(x) \Rightarrow ||D_x f(v)|| \leqslant a||v||. \tag{18}$$

There are other definitions of hyperbolicity and related quantities such as non-uniform partial hyperbolicity and dominated splittings; more specific characteristics and definitions can be found in [7, 19, 20, 58, 59]. The key provided by definition 7 is the allowance of centre bundles, zero Lyapunov exponents and, in general, neutral directions, which are not allowed in strict hyperbolicity. Thus, we are allowed to keep the general framework of good topological structure, but we lose structural stability. With non-trivial partial hyperbolicity (i.e. E^c is not null), stable ergodicity replaces structural stability as the notion of dynamic stability in the Pugh–Shub stability conjecture (conjecture (5) of [60]). Thus, what is left is to again attempt to show the extent to which stable ergodicity persists.

In numerical simulations we will never observe an orbit on the unstable, stable or centre manifolds (or the bundles). Thus, we will need a global notion of stability averaged along a given orbit (which will exist under weak ergodic assumptions). The notion we seek is captured by the spectrum of Lyapunov exponents.

[6] If E^c is trivial, f is simply Anosov, or strictly hyperbolic.

Definition 8 (Lyapunov Exponents). *Let $f : M \to M$ be a diffeomorphism (i.e. discrete time map) on a compact Riemannian manifold of dimension m. Let $|\cdot|$ be the norm on the tangent vectors induced by the Riemannian metric on M. For every $x \in M$ and $v \in T_x M$ Lyapunov exponent at x is denoted:*

$$\chi(x, v) = \limsup_{t\to\infty} \frac{1}{t} \log ||Df^n v||. \tag{19}$$

Assume the function $\chi(x, \cdot)$ has only finitely many values on $T_x M$ {0} (this assumption may not be true for our dynamical systems) which we denote $\chi_1^f(x) < \chi_2^f(x) \ldots < \chi_m^f(x)$. Next denote the filtration of $T_x M$ associated with $\chi(x, \cdot)$, $\{0\} = V_0(x) \subsetneqq V_1(x) \subsetneqq \ldots \subsetneqq V_m(x) = T_x M$, where $V_i(x) = \{v \in T_x M | \chi(x, v) \leqslant \chi_i(x)\}$. The number $k_i = \dim(V_i(x)) - \dim(V_{i-1}(x))$ is the multiplicity of the exponent $\chi_i(x)$. In general, for our networks over the parameter range we are considering, $k_i = 1$ for all $0 < i \leqslant m$. Given the above, the Lyapunov spectrum for f at x is defined as:

$$\mathrm{Sp}\chi(x) = \{\chi_j^k(x) | 1 \leqslant i \leqslant m\}. \tag{20}$$

(For more information regarding Lyapunov exponents and spectra see [61–64]).

A more computationally motivated formula for the Lyapunov exponents is given as

$$\chi_j = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \ln(\langle (Df_k \cdot \delta x_j)^T, (Df_k \cdot \delta x_j)\rangle) \tag{21}$$

where $\langle , \rangle$ is the standard inner product, δx_j is the jth component of the x variation[7] and Df_k is the 'orthogonalized' Jacobian of f at the kth iterate of $f(x)$. Through the course of our discussions we will dissect equation (21) further. It should also be noted that Lyapunov exponents have been shown to be independent of coordinate system. Thus, the specifics of our above definition do not affect the outcome of the exponents.

For the systems we study, there could be an x dependence on equations (19) and (21). As will be seen in later sections, we do not observe much of an x-dependence on the Lyapunov exponents over the parameter ranges considered. The existence (or lack) of multiple attractors has been partially addressed in a conjecture in [65]; however, a more systematic study is currently under way.

In general, the existence of Lyapunov exponents is established by a multiplicative ergodic theorem (for a nice example, see theorem (1.6) in [66]). There exist many such theorems for various circumstances. The first multiplicative ergodic theorem was proved by Oseledec [67]; many others—[20, 68–72]—have subsequently generalized his original result. We will refrain from stating a specific multiplicative ergodic theorem; the conditions necessary for the existence of Lyapunov exponents are exactly the conditions we place on our function space in section 2.3, that is, a C^r ($r > 0$) map of a compact manifold M (or open set U) to itself and an f-invariant probability measure ρ, on M (or U). For specific treatments we leave the curious reader to study the aforementioned references, noting that our construction follows from [20, 59, 69].

There is an intimate relationship between Lyapunov exponents and global stable and unstable manifolds. In fact, each Lyapunov exponent corresponds to a global manifold, and often the existence of the various global manifolds is explicitly tied to the existence of the respective Lyapunov exponents. We will be using the global manifold structure as our measure of topological equivalence and the Lyapunov exponents to classify this global structure. Positive Lyapunov exponents correspond to global unstable manifolds, and negative Lyapunov exponents correspond to global stable manifolds. We will again refrain from stating

[7] In a practical sense, the x variation is the initial separation or perturbation of x.

the existence theorems for these global manifolds (see [64] for a good discussion of this) and instead note that in addition to the requirements for the existence of Lyapunov exponents, the existence of global stable/unstable manifolds corresponding to the negative/positive Lyapunov exponents requires Df to be injective. For specific global unstable/stable manifold theorems see [69].

Finally, the current best solution to the structural stability conjecture of Palis and Smale, the result that links hyperbolicity to structural stability, is captured in the following theorem.

Theorem 2 (Mañé [5] theorem A, Robbin [3], Robinson [73]). *A C^1 diffeomorphism (on a compact, boundaryless manifold) is structurally stable if and only if it satisfies axiom A and the strong transversality condition.*

Recall that axiom A says the diffeomorphism is hyperbolic with dense periodic points on its non-wandering set Ω ($p \in \Omega$ is non-wandering if for any neighbourhood U of x, there is an $n > 0$ such that $f^n(U) \cap U \neq 0$). We will save a further explicit discussion of this interrelationship for a later section, noting that much of this report investigates the above notions and how they apply to our set of maps. Finally, for a nice, sophisticated introduction to the above topics see [62] or [74].

2.3. *Conditions needed for the existence and computation of Lyapunov exponents*

We will follow the standard constructions for the existence and computation of Lyapunov exponents as defined by the theories of Katok [68], Ruelle [66, 69], Pesin [70–72], Brin and Pesin [20] and Burns *et al* [59].

Let $\mathcal{H}$ be a separable real Hilbert space (for practical purposes R^n) and let X be an open subset of $\mathcal{H}$. Next let (X, Σ, ρ) be a probability space where Σ is a σ-algebra of sets and ρ is a probability measure, $\rho(X) = 1$ (see [75] for more information). Now consider a C^r $(r > 1)$ map $f_t : X \mapsto X$ which preserves ρ (ρ is f-invariant, at least on the unstable manifolds) defined for $t \geqslant T_0 \geqslant 0$ such that $f_{t_1+t_2} = f_{t_1} \circ f_{t_2}$ and that $(x, t) \mapsto f_t(x)$, $Df_t(x)$ is continuous from $X \times [T_0, \infty)$ to X and bounded on $\mathcal{H}$. Assume that f has a compact invariant set

$$\Lambda = \left\{ \bigcap_{t > T_0} f_t(X) | f_t(\Lambda) \subseteq \Lambda \right\} \tag{22}$$

and Df_t is a compact bounded operator for $x \in \Lambda$, $t > T_0$. Finally, endow f_t with a scalar parameter $s \in [0 : \infty]$. This gives us the space (a metric space—the metric will be defined heuristically in section 2.1.4) of one parameter, C^r measure-preserving maps from bounded compact sets to themselves with bounded first derivatives. It is for a space of the above mappings that Ruelle shows the existence of Lyapunov exponents [69]; similar requirements are made by Brin and Pesin [20] in a slightly more general setting. The systems we will study are dissipative dynamical systems, and thus area-contracting, so ρ will not be absolutely continuous with respect to Lebesgue measure on X. However, to compute Lyapunov exponents, it is enough for there to exist invariant measures that are absolutely continuous with respect to Lebesgue on the unstable manifolds [64, 76–78]. Thus, the Lyapunov exponents in the systems we study are computed relative to SRB measures [64] that are assumed, but not proved, to exist for systems we study. Implications of invariant measures for dissipative dynamical systems such as those studied here can be found in [79].

We can take X in the above construction to be the R^d of section 2.1.1. The neural networks we use map their domains to compact sets; moreover, because they are constructed as time-delays, their domains are also compact. Further, their derivatives are bounded up to arbitrary order; although for our purposes, only the first order need be bounded. Because the neural networks are deterministic and bounded, there will exist an invariant set of some type. We are relegated to assuming the existence of SRB measures with which we can calculate the Lyapunov exponents because proving the existence of SRB measures, even for relatively simple dissipative dynamical systems, is non-trivial [80, 81]. Indeed, there remains much work to achieve a full understanding of Lyapunov exponents for general dissipative dynamical systems that are not absolutely continuous; for a current treatment see [61] or [64]. The specific measure theoretic properties of our networks (i.e. issues such as absolute continuity, uniform/non-uniform hyperbolicity, basin structures) is a topic of current investigation. In general, up to the accuracy, interval of initial conditions and dimension we are concerned with in this paper, the networks observed do not have multiple attractors and thus have a single invariant measure. We will not prove this here and, in fact, know of counter-examples to the former statement for some parameter settings.

2.4. Definitions for numerical arguments

Because we are conducting a numerical experiment, it is necessary to present notions that allow us to test our conjectures numerically. We will begin with a notion of continuity. The heart of continuity is based on the following idea: if a neighbourhood about a point in the domain is shrunk, this implies a shrinking of a neighbourhood of the range. However, we do not have infinitesimals at our disposal. Thus, our statements of numerical continuity will necessarily have a statement regarding the limits of numerical resolution below which our results are uncertain.

Let us now begin with a definition of bounds on the domain and range:

Definition 9 (ϵ_{num}). ϵ_{num} *is the numerical accuracy of a Lyapunov exponent,* χ_j.

Definition 10 (δ_{num}). δ_{num} *is the numerical accuracy of a given parameter under variation.*

Now, with our ϵ_{num} and δ_{num} defined as our numerical limits in precision, let us define numerical continuity of Lyapunov exponents.

Definition 11 (num-continuous Lyapunov exponents). *Given a one parameter map* $f : R^1 \times R^d \to R^d$, $f \in C^r$, $r > 0$, *for which characteristic exponents* χ_j *exist (a.e. with respect to an SRB measure). The map* f *is said to have num-continuous Lyapunov exponents at* $(\mu, x) \in R^1 \times R^d$ *if for* $\epsilon_{\mathrm{num}} > 0$ *there exists a* $\delta_{\mathrm{num}} > 0$ *such that if*

$$|s - s'| < \delta_{\mathrm{num}} \tag{23}$$

then

$$|\chi_j(s) - \chi_j(s')| < \epsilon_{\mathrm{num}} \tag{24}$$

for $s, s' \in R^1$, *for all* $j \in N$ *such that* $0 < j \leqslant d$.

Another useful definition related to continuity is that of a function being Lipschitz continuous.

Definition 12 (num-Lipschitz). *Given a one parameter map* $f : R^1 \times R^d \to R^d$, $f \in C^r$, $r > 0$, *for which characteristic exponents* χ_j *exist (and are the same under all invariant measures), the map* f *is said to have num-Lipschitz Lyapunov exponents at* $(\mu, x) \in R^1 \times R^d$ *if there exists a real constant* $0 < k_{\chi_j}$ *such that*

$$|\chi_j(s) - \chi_j(s')| < k_{\chi_j}|s - s'|. \tag{25}$$

Further, if the constant $k_{\chi_j} < 1$*, the Lyapunov exponent is said to be contracting*[8] *on the interval* $[s, s']$ *for all* s' *such that* $|s - s'| < \delta_{\text{num}}$.

Note that neither of these definitions imply strict continuity, but rather they provide bounds on the difference between the change in parameter and the change in Lyapunov exponents. It is important to note that these notions are highly localized with respect to the domain in consideration. We will not imply some sort of global continuity using the above definitions; rather, we will use these notions to imply that Lyapunov exponents will continuously (within numerical resolution) cross through zero upon parameter variation. We can never numerically prove that Lyapunov exponents do not jump across zero, but for most computational exercises, a jump across zero that is below numerical precision is not relevant. This notion of continuity will aid in arguments regarding the existence of periodic windows in parameter space.

Let us next define a Lyapunov exponent zero-crossing.

Definition 13 (Lyapunov exponent zero-crossing). *A Lyapunov exponent zero-crossing is simply the point* s_{χ_j} *in parameter space such that a Lyapunov exponent continuously (or num-continuously) crosses zero, e.g. for* $s - \delta$, $\chi_i > 0$*, and for* $s + \delta$, $\chi_i < 0$.

For this report, a Lyapunov exponent zero-crossing is a transverse intersection with the real line. For our networks non-transversal intersections of the Lyapunov exponents with the real line certainly occur, but for the portion of parameter space we are investigating, they are extremely rare. Along the route-to-chaos for our networks, such non-transversal intersections are common, but we will save the discussion of that topic for a different report. Orbits for which the Lyapunov spectrum can be defined (in a numerical sense, Lyapunov exponents are defined when they are convergent), yet at least one of the exponents is zero, are called non-trivially num-partially hyperbolic. We must be careful making statements with respect to the existence of zero Lyapunov exponents implying the existence of centre manifolds E^c because zero exponents may not correspond to a manifold.

Lastly, we define a notion of denseness for a numerical context. There are several ways of achieving such a notion—we will use the notion of a dense sequence.

Definition 14 (ϵ-dense). *Given an* $\epsilon > 0$*, an open interval* $(a, b) \subset R$*, and a sequence* $\{c_1, \ldots, c_n\}$, $\{c_1, \ldots, c_n\}$ *is* ϵ*-dense in* (a, b) *if there exists an* n *such that for any* $x \in (a, b)$*, there is an* i, $1 \leqslant i < n$*, such that* $\mathrm{dist}(x, c_i) < \epsilon$.

In reality however, we will be interested in a sequence of sequences that are 'increasingly' ϵ-dense in an interval (a, b). In other words, for the sequence of sequences

$$\begin{matrix} c_1^1, & \cdots, & c_{n_1}^1 \\ c_1^2, & \cdots, & c_{n_2}^2 \\ \vdots & \vdots & \vdots \end{matrix}$$

where $n_{i+1} > n_i$ (i.e. for a sequence of sequences with increasing cardinality), the subsequent sequences for increasing n_i become a closer approximation of an ϵ-dense sequence.

Definition 15 (Asymptotically dense (a-dense)). *A sequence* $S_j = \{c_1^j, \ldots, c_{n_j}^j\} \subset (a, b)$ *of finite subsets is asymptotically dense in* (a, b)*, if for any* $\epsilon > 0$*, there is an* N *such that* S_j *is* ϵ*-dense if* $j \geqslant N$.

[8] Note, there is an important difference between the Lyapunov exponent contracting, which implies some sort of convergence to a particular value, versus a negative Lyapunov exponent that implies a contracting direction on the manifold or in phase space.

For an intuitive example of this, consider a sequence S of k numbers where $q_k \in S, q_k \in (0, 1)$. Now increase the cardinality of the set, spreading elements in such a way that they are uniformly distributed over the interval. Density is achieved with the cardinality of infinity, but clearly, with a finite but arbitrarily large number of elements, we can achieve any approximation to a dense set that we wish. There are, of course, many ways we can have a countably infinite set that is not dense, and, as we are working with numerics, we must concern ourselves with how we will approach this asymptotic density. We now need a clear understanding of when this definition will apply to a given set. There are many pitfalls; for instance, we wish to avoid sequences such as $(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots)$. We will, in the section that addresses a-density, state the necessary conditions for an a-dense set for our purposes.

3. Conjectures

The point of this exercise is to begin to specify, in a computational framework, the variation in Lyapunov exponents and hyperbolicity in a space of high-dimensional dynamical systems. To achieve this goal, we will make four statements or conjectures that we will subsequently support with numerical evidence. To begin, let us specify a condition that will be very valuable to the formulation of our ideas.

Condition 1. *Given a map (neural network) as defined in section 2.1.4, if the parameter $s \in R^1$ is varied num-continuously, then the Lyapunov exponents vary num-continuously.*

Since there are many counterexamples to this condition, many of our results rely upon our ability to show how generally this condition applies in high-dimensional systems.

Definition 16 (Chain link set). *Assume f is a mapping (neural network) as defined in section 2.1.4. A* ***chain link set*** *is denoted as*

$$V = \{s \in R \mid \chi_j(s) \neq 0 \text{ for all } 0 < j \leqslant d \text{ and } \chi_j(s) > 0 \text{ for some } j > 0\}.$$

If $\chi_j(s)$ is continuous at its Lyapunov exponent zero-crossing, as we will show later (condition (1)), then V is open. Next, let C_k be a connected component of the closure of V, $\overline{V}$. It can be shown that $C_k \cap V$ is a union of disjoint, adjacent open intervals of the form $\bigcup_i (a_i, a_{i+1})$.

Definition 17 (Bifurcation link set). *Assume f is a mapping (neural network) as defined in section 2.1.4. Denote a* ***bifurcation link set*** *of $C_k \cap V$ as*

$$V_i = (a_i, a_{i+1}). \tag{26}$$

Assume the number of positive Lyapunov exponents for each $V_i \subset V$ remains constant. If, upon a monotonically increasing variation in the parameter s, the number of positive Lyapunov exponents for V_i is greater than the number of positive Lyapunov exponents for V_{i+1}, V is said to be LCE decreasing. Specifically, the endpoints of the V_i are the points where there exist Lyapunov exponent zero crossings. We are not particularly interested in these sets; however, we are interested in the collection of endpoints adjoining these sets.

Definition 18 (Bifurcation chain subset). *Let V be a chain link set, and C_k a connected component of $\overline{V}$. A* ***bifurcation chain subset*** *of $C_k \cap V$ is denoted as*

$$U_k = \{a_i\} \tag{27}$$

or equivalently

$$U_k = \partial(C_k \cap V). \tag{28}$$

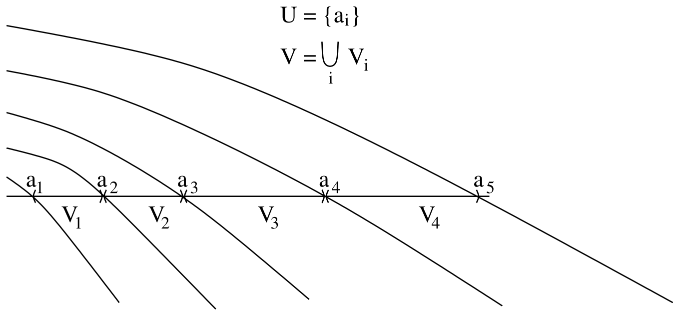

Figure 2. An intuitive diagram for chain link sets, V, bifurcation link sets, V_i, and bifurcation chain sets, U, for an LCE decreasing chain link set V.

For the purposes of this work, we will consider a bifurcation chain subset U such that a_1 corresponds to the last zero crossing of the least positive exponent and a_n will depend upon the specific case and dimension. For comparison with later figures $a_1 \sim 0.5$ and $a_n \sim 6$; in higher-dimensional networks, $a_n \sim 6$ will correspond to a much higher n than for a low-dimensional network. For an intuitive picture of what we wish to depict with the above definitions, consider figure 2.

We can now state the conjectures and an outline of what we will test and why those tests will verify our claims.

Conjecture 1 (Hyperbolicity violation). *Assume f is a mapping (neural network) as defined in section 2.1.4 with a sufficiently large number of dimensions, d. There exists at least one bifurcation chain subset U.*

The intuition arises from a straightforward consideration of the neural network construction in section 2.1.4. From the consideration of our specific neural networks and their activation function, tanh(), it is clear that variation of the scaling parameter, s, on the variance of the interaction weights ω forces the neural networks from a linear region through a non-linear region and into a binary region. This implies that, given a neural network that is chaotic for some value of s, upon the monotonically increasing variation of s from zero, the dynamical behaviour will begin at a fixed point, proceed through a sequence of bifurcations, become chaotic and eventually become periodic. If the number of positive Lyapunov exponents can be shown to increase with the dimension of the network and if the Lyapunov exponents can be shown to vary relatively continuously with respect to parameter variation with increasing dimension then there will be many points along the parametrized curve such that there will exist neutral directions. The ideas listed above provide the framework for computational verification of conjecture (2). We must investigate conjecture (1) with respect to the subset U becoming *a-dense* in its closure and the existence of very few (ideally a single) connected components of $\overline{V}$.

Conjecture 2 (Existence of a codimension ϵ bifurcation set). *Assume f is a mapping (neural network) as defined in section 2.1.4 with a sufficiently large number of dimensions, d, and a bifurcation chain set U as per conjecture (1). The two following statements hold and are equivalent:*

(i) In the infinite-dimensional limit, the cardinality of U will go to infinity, and the length $\max |a_{i+1} - a_i|$ *for all i will tend to zero on a one-dimensional interval in parameter space. In other words, the bifurcation chain set U will be a-dense in its closure, $\overline{U}$.*

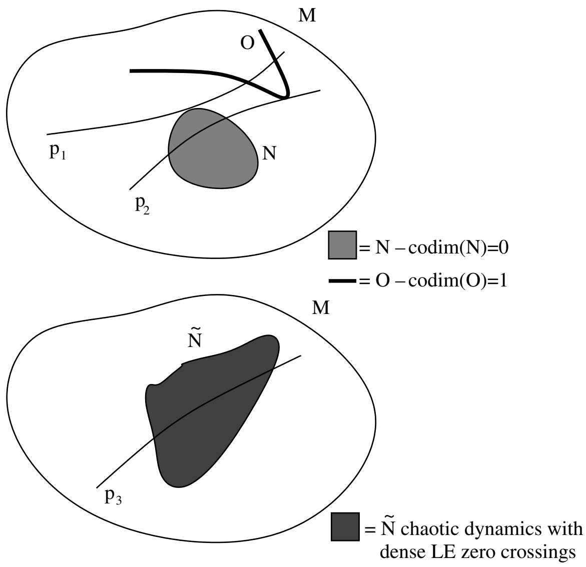

Figure 3. The top drawing represents various standard pictures from transversality theory. The bottom drawing represents an idealized version (in higher dimensions) of transversality catering to our arguments.

(ii) In the asymptotic limit of high dimension, for all $s \in U$ and for all f at s, an arbitrarily small perturbation δ_s of s will produce a topological change. The topological change will correspond to a different number of global stable and unstable manifolds for f at s compared with f at $s + \delta$.

Assume M is a C^r manifold of topological dimension d, and N is a submanifold of M. The codimension of N in M is defined by $\text{codim}(N) = \dim(M) - \dim(N)$. If there exists a curve p through M such that p is transverse to N and $\text{codim}(N) \leqslant 1$, then there will not exist an arbitrarily small perturbation to p such that p will become non-transverse to N. Moreover, if $\text{codim}(N) = 0$ and $p \bigcap N \subset \text{int}(N)$ then there does not even exist an arbitrarily small perturbation of p such that p intersects N at a single point of N, i.e. the intersection cannot be made non-transverse with an arbitrarily small perturbation.

The former paragraph can be more easily understood from figure 3 where we have drawn four different circumstances. The first circumstance, the curve $p_1 \cap N$, is an example of a non-transversal intersection with a codimension 0 submanifold. This intersection can be perturbed away with an arbitrarily small perturbation of p_1. The intersection, $p_2 \cap N$, is a transversal intersection with a codimension 0 submanifold, and this intersection cannot be perturbed away with an arbitrarily small perturbation of p_2. Likewise, the intersection, $p_1 \cap O$, which is an example of a transversal intersection with a codimension 1 submanifold, cannot be made non-transverse or null by an arbitrarily small perturbation of p_1. The intersection $p_2 \cap O$ is a non-transversal intersection with a codimension 1 submanifold and can be perturbed away with an arbitrarily small perturbation of p_2. This outlines the avoidability of codimension 0 and 1 submanifolds with respect to curves through the ambient manifold M. The point is that non-null, transversal intersections of curves with codimension 0 or 1 submanifolds cannot be made non-transversal with arbitrarily small perturbations of the curve. Transversal intersections of curves with codimension 2 submanifolds, however, can always be removed by an arbitrarily small perturbation due to the existence of a 'free' dimension. A practical example of such would be the intersection of a curve with another curve in R^3—one can always pull apart the two curves simply by 'lifting' them apart.

In the circumstance proposed in conjecture (2), the set U ($\tilde{N}$ in figure 3) will always have codimension d because U consists of finitely many points; thus, any intersection with U can be removed by an arbitrarily small perturbation. The point is that, as U becomes a-dense in $\bar{U}$, $p_3 \bigcap \bar{U} = 0$ becomes more and more unlikely, and the perturbations required to remove the intersections of p_3 with U (again, $\tilde{N}$ as in figure 3) will become more and more bizarre. For a low-dimensional example, think of a ball of radius r in R^3 that is populated by a finite set of evenly distributed points, denoted S_i, where i is the number of elements in S_i. Next fit a curve p through that ball in such a way that p does not hit any points in S_i. Now, as the cardinality of S_i becomes large, if S_i is a-dense in the ball of radius r, for the intersection of p with S_i to remain null, p will need to become increasingly kinky. Moreover, continuous, linear transformations of p will become increasingly unlikely to preserve $p \cap S_i = 0$. It is this type of behaviour with respect to parameter variation that we are arguing for with conjecture (2). However, figure 3 should only be used as an tool for intuition—our conjectures are with respect to a particular interval in parameter space and not a general curve in parameter space, let alone a family of curves or a high-dimensional surface. Conjecture (2) is a first step towards a more complete argument with respect to the above scenario. For more information concerning the origin of the above picture, see [82] or [83]. With the development of more mathematical language, it is likely that this conjecture can be specified with the notion of prevalence.

To understand roughly why we believe conjecture (2) is reasonable, first take condition (1) for granted (we will expend some effort showing where condition (1) is reasonable). Next assume there are arbitrarily many Lyapunov exponents near 0 along some interval of parameter space and that the Lyapunov exponent zero-crossings can be shown to be a-dense with increasing dimension. Further, assume that on the aforementioned interval, V is LCE decreasing. Since varying the parameters continuously on some small interval will move Lyapunov exponents continuously, small changes in the parameters will guarantee a continual change in the number of positive Lyapunov exponents. One might think of this intuitively relative to the parameter space as the set of Lyapunov exponent zero-crossings forming a codimension 0 submanifold with respect to the particular interval of parameter space. However, we will never achieve such a situation in a rigorous way. Rather, we will have an a-dense bifurcation chain set U, which will have codimension 1 in R with respect to topological dimension. As the dimension of f is increased, U will behave more like a codimension 0 submanifold of R. Hence the metaphoric language, codimension ϵ bifurcation set. The set U will always be a codimension one submanifold because it is a finite set of points. Nevertheless, if U tends towards being dense in its closure, it will behave increasingly like a codimension zero submanifold. This argument will not work for the entirety of the parameter space, and thus we will show where, to what extent, and under what conditions U exists and how it behaves as the dimension of the network is increased.

Conjecture 3 (Periodic window probability decreasing). *Assume f is a mapping (neural network) as defined in section 2.1.4 and a bifurcation chain set U as per conjecture (1). In the asymptotic limit of high dimension, the length of the bifurcation chain sets, $l = |a_n - a_1|$, increases such that the cardinality of $U \to m$ where m is the maximum number of positive Lyapunov exponents for f. In other words, there will exist an interval in parameter space (e.g. $s \in (a_1, a_n) \sim (0.1, 4)$) where the probability of the existence of a periodic window will go to zero (with respect to Lebesgue measure on the interval) as the dimension becomes large.*

This conjecture is somewhat difficult to test for a specific function since adding inputs completely changes the function. Thus, the curve through the function space is an abstraction which is not afforded by our construction. We will save a more complete analysis (e.g. a search for periodic windows along a high-dimensional surface in parameter space) of conjecture (3) for a different report. In this work, conjecture (3) addresses a very practical matter, for it implies the existence of a much smaller number of bifurcation chain sets. The previous conjectures allow for the existence of many of these bifurcation chains sets, U, separated by windows of periodicity in parameter space. However, if these windows of periodic dynamics in parameter space vanish, we could end up with only one bifurcation chain set—the ideal situation for our arguments. We will not claim such; however, we will claim that the length of the set U we are concerning ourselves with in a practical sense will increase with increasing dimension, largely due to the disappearance of periodic windows on the closure of V. With respect to this report, all that needs to be shown is that the window sizes along the path in parameter space for a variety of neural networks decreases with increasing dimension. From a qualitative analysis, it will be somewhat clear that the above conjecture is reasonable.

If this paper were actually making statements we could rigorously prove; conjectures (1), (2) and (3) would function as lemmas for conjecture (4).

Conjecture 4. *Assume f is a mapping (neural network) as defined in section 2.1.4 with a sufficiently large number of dimensions, d, a bifurcation chain set U as per conjecture (1) and the chain link set V. The perturbation size δ_s of $s \in C_{max}$, where C_{max} is the largest connected component of $\overline{V}$, for which $f|_{C_k}$ remains structurally stable, goes to zero as $d \to \infty$.*

Specific cases and the lack of density of structural stability in certain sets of dynamical systems has been proved long ago. These examples were, however, very specialized and carefully constructed circumstances and do not speak of the commonality of structural stability failure. Along the road to investigating conjecture (4) we will show that structural stability will not, in a practical sense, be observable for a large set of very high-dimensional dynamical systems along certain, important intervals in parameter space even though structural stability is a property that will exist on that interval with probability one (with respect to Lebesgue measure). This conjecture might appear to contradict some well-known results in stability theory. A careful analysis of this conjecture, and its relation to known results, will be discussed in sections 7.1.4 and 7.3.1.

The larger question that remains, however, is whether conjecture (4) is valid on high-dimensional surfaces in parameter space. We believe this is a much more difficult question with a much more complicated answer. A highly related problem of whether chaos persists in high-dimensional dynamical systems is partially addressed in [65].

4. Numerical errors and Lyapunov exponent calculation

The Lyapunov exponents are calculated by equation (21) using the well-known and standard techniques given in [41, 84–86]. We employ the modified Gram–Schmidt orthogonalization technique because it is more numerically stable than the standard Gram–Schmidt technique. Moreover, the orthonormalization is computed at every time step. The results are not altered when we use other orthonormalization algorithms (see [87] for a survey of different techniques). The explicit calculation of LCEs in the context we utilize is discussed well in [88]. Numerical error analysis for the Lypaunov exponent calculation has been studied in various contexts; see [89] or [90] for details. Determining errors in the Lyapunov exponents is not an exact science; for our networks such errors vary a great deal in different regions in s space. For instance, near the first bifurcation from a fixed point can require up to 100 000 or more iterations to converge to an attractor and 500 00 more iterations for the Lyapunov spectrum to converge. The region of parameter space we study here has particularly fast convergence.

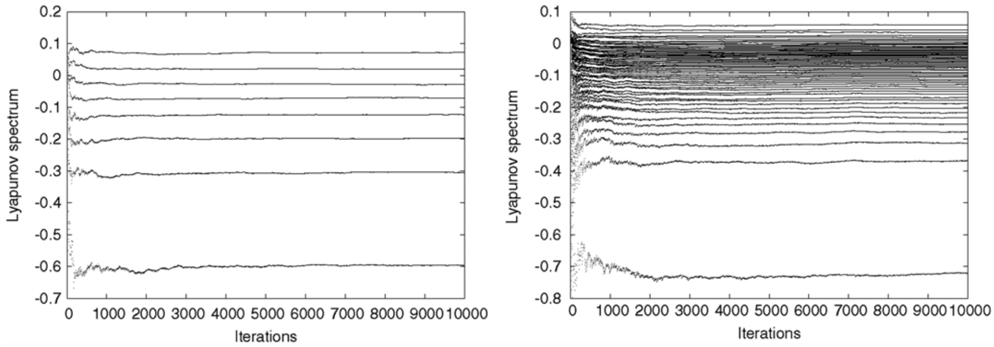

Figure 4. LE spectrum versus iteration for individual networks with 32 neurons and 16 (left, only the largest 8 are shown) and 64 (right) dimensions.

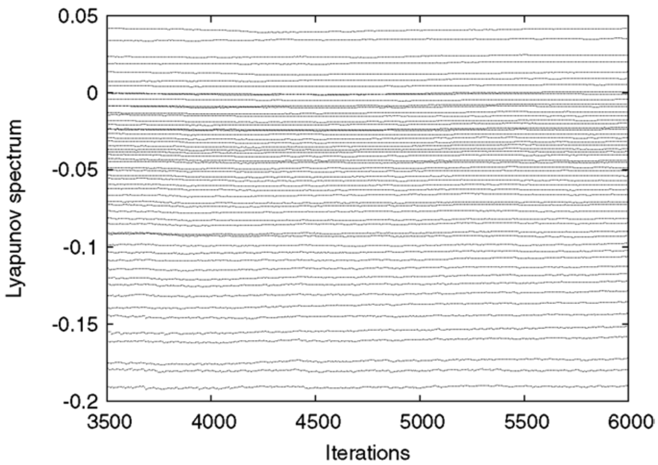

Figure 5. Close-up of LE spectrum versus iteration: 32 neurons, 64 dimensions.

Consider figure 4, depicting plots of the Lyapunov spectrum versus the first 10000 iterations for two networks with 16 and 64 dimensions. After approximately 3000 time steps, all the large transients have essentially vanished, and aside from a slight variation (especially on a time-scale that is long compared with a single time-step) the exponents appear to have converged. For the case with 16 dimensions, the exponents also appear to have converged. The resolution for the network with 64 dimensions is not fine enough to verify a distinction between exponents; thus, consideration of figure 5 demonstrates clearly that the exponents converge within the inherent errors in the calculation and are entirely distinct for greater than 5500 time steps. It is worth noting that there are times when very long term transients occur in our networks. These transients would not be detectable from the figures we have presented, but these problem cases usually exist only near bifurcation points. For the cases we are considering, these convergence issues do not seem to affect our results[9] (see chapter 10 of [35] for more information).

[9] When an exponent is very nearly zero it can tend to fluctuate above and below zero, but it is always very near zero. Thus, although it might be difficult to resolve zero exactly—which is to be expected—the exponent is clearly very near zero which is all that really matters for our purposes.

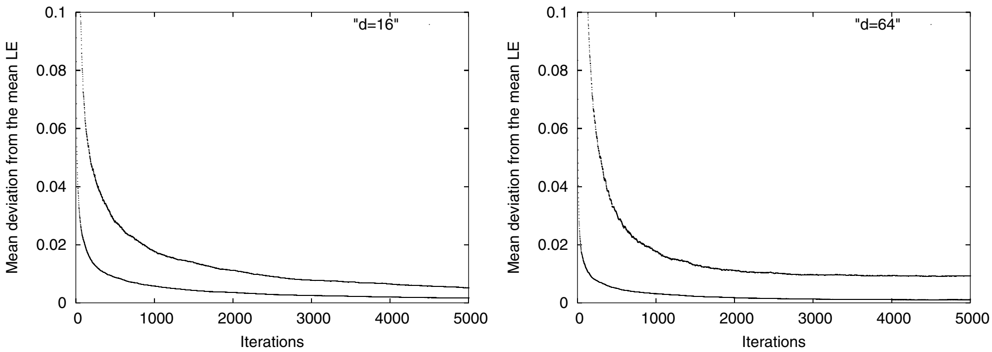

Figure 6. Mean deviation from the mean of the largest and most negative Lyapunov exponent per time-step for an ensemble of 1000 networks with 32 neurons and 16 (left) and 64 (right) dimensions.

Figures 4 and 5 provide insight into how the individual exponents for individual networks converge; we now must establish the convergence of the Lyapunov exponents for a large set of neural networks and present a general idea of the numerical variance (ϵ_m) in the Lyapunov exponents. We will achieve this by calculating the Lyapunov spectrum for an individual network for 5000 time steps, calculating the mean of each exponent in the spectrum, for each time step calculating the deviation of the exponent from the mean of that exponent and repeating the procedure for 1000 networks taking the mean of the deviation from the mean exponent at each time step. Figure 6 represents the analysis in the former statement. This figure demonstrates clearly that the deviation from the mean exponent, even for the most negative exponent (which has the largest error), drops below 0.01 after 3000 time steps. The fluctuations in the largest Lyapunov exponent lie in the 10^{-3} range for 3000 time steps. Figure 6 also substantiates three notions: a measurement of how little the average exponent strays from its mean value; a measurement of the similarity of this characteristic over the ensemble of networks and a general intuition with respect to the accuracy of our exponents, $\epsilon_m < 0.01$ for 5000 time steps.

5. Numerical arguments for preliminaries

Before we present our arguments supporting our conjectures we must present various preliminary results. Specifically, the num-continuity of the Lyapuonv exponents, the a-density of Lyapunov exponent zero-crossings, and the existence of arbitrarily many positive exponents given an arbitrarily large dimension.

5.1. num-continuity

Testing the num-continuity of Lyapunov exponents will be two-fold. First, we will need to investigate, for a specific network f, the behaviour of Lyapunov exponents versus variation of parameters. Second, indirect, yet strong evidence of the num-continuity will also come from investigating how periodic window size varies with dimension and parameters. It is important to note that when we refer to continuity, we are referring to a very local notion of continuity. Continuity is always in reference to the set upon which something (a function, a mapping, etc) is continuous. In the analysis below, the neighbourhoods upon which continuity of the Lyapunov exponents are examined are plus and minus one parameter increment. This is all

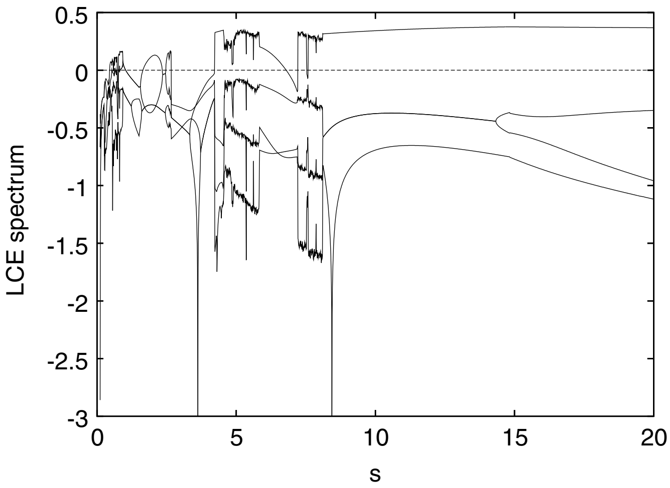

Figure 7. LE spectrum: 32 neurons, 4 dimensions, $s \in [0 : 20]$.

that is necessary for our purposes, but this analysis cannot guarantee strict continuity along, say, $s \in [0.1, 10]$, but rather continuity along little linked bits of the interval $[0.1, 10]$.

5.1.1. Qualitative analysis. Qualitatively, our intuition for num-continuity comes from examining hundreds of Lyapunov spectrum plots versus parameter variation. In this vein, figures 7 and 8 present the difference between low and higher dimensional Lyapunov spectra.

In figure 8, the Lyapunov exponents look continuous within numerical errors (usually ± 0.005). Figure 8 by itself provides little more than an intuitive picture of what we are arguing. As we will be making arguments that the Lyapunov spectrum will become smoother, periodic windows will disappear, and so on, with increasing dimension, figure 7 shows a typical graph of the Lyapunov spectrum versus parameter variation for a neural network with 32 neurons and 4 dimensions. The contrast between figures 8 and 7 intuitively demonstrates the increase in continuity we are claiming.

Although figures 7 and 8 shows that, as the dimension is increased, the Lyapunov exponents appear to be a more continuous function of s, the figures alone do not verify num-continuity. In fact, note that pathological discontinuities have been observed in networks with as many as 32 dimensions. The existence of pathologies for higher dimensions is not a problem we are prepared to answer in depth; it can be confidently said that as the dimension (number of inputs) is increased, the pathologies appear to become vanishingly rare (this is noted over our observation of several thousand networks with dimensions ranging from 4 to 256).

5.1.2. Quantitative and numerical analysis. Begin by considering the num-continuity of two particular networks while varying the s parameter. Figure 9 depicts the mean difference in each exponent between parameter values summed over all the exponents. The parameter increment is $\delta s = 0.01$. The region of particular interest is between $s = 0$ and 6. Considering this range, it is clear that the variation in the mean of the exponents versus variation in s decreases with dimension. The 4-dimensional network not only has a higher baseline of num-continuity but it also has many large spikes that correspond to the appearance of large qualitative changes in the dynamics. As the dimension is increased, considering the 64-dimensional case, the baseline of num-continuity is decreased, and the large spikes disappear. The spikes in the 4-dimensional

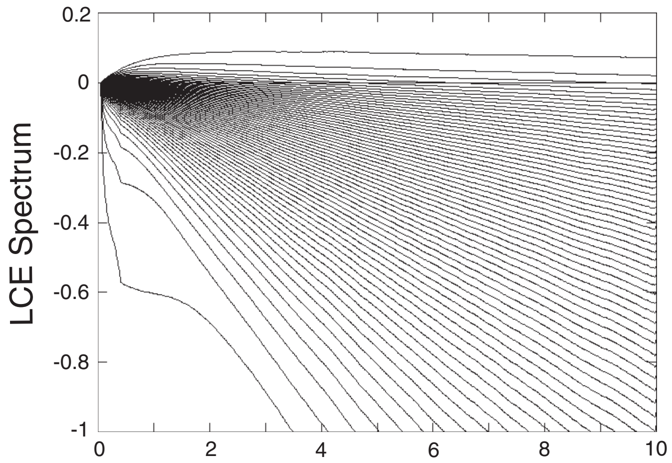

Figure 8. LE spectrum: 32 neurons, 64 dimensions, $s \in [0.11, 10]$.

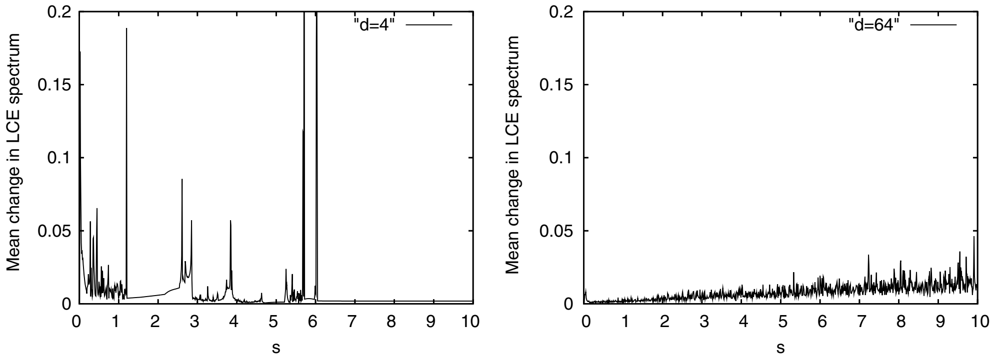

Figure 9. num-continuity (mean of $|\chi_i(s) - \chi_i(s + \delta s)|$ for each i) versus parameter variation: 32 neurons, 4 (left) and 64 (right) dimensions.

case can be directly linked to the existence of periodic windows and bifurcations that result in a dramatic topological change. This is one verification of num-continuity of Lyapunov exponents. These two cases are typical, but it is clear that the above analysis, although persuasive, is not adequate for our needs. We will thus resort to a statistical study of the above plots.

The statistical support we have for our claim of increased num-continuity will again focus on the parameter region between $s = 0$ and 6, the region in parameter space near the maxima of entropy, Kaplan–Yorke dimension and the number of positive Lyapunov exponents. Figure 10 considers the num-continuity for parameter values of 0–6. The points on the plot correspond to the mean (for a few hundred networks) of the mean exponent change between parameter values or

$$\mu^d = \frac{1}{Z} \sum_{k=1}^{Z} \frac{\sum_{i=1}^{d} |\chi_i^k(s) - \chi_i^k(s + \delta s)|}{d}, \tag{29}$$

where Z is the total number of networks of a given dimension. Figure 10 clearly shows that as the dimension is increased, for the same computation time, both the mean exponent change versus the parameter variation per network and the standard deviation of the exponent change

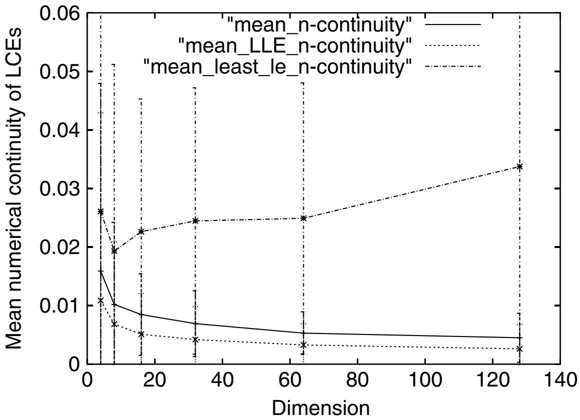

Figure 10. Mean num-continuity, num-continuity of the largest and the most negative Lyapunov exponent of many networks versus their dimension. The error bars are the standard deviation about the mean for the networks considered.

decrease substantially as the dimension is increased[10]. Of course the mean change over all the exponents allows for the possibility for one exponent (possibly the largest exponent) to undergo a relatively large change while the other exponents change very little. For this reason, we have included the num-continuity of the largest and the most negative exponents versus parameter change. The num-continuity of the largest exponents is very good, displaying a small standard deviation across many networks. The error in the most negative exponent is inherent to our numerical techniques (specifically the Gram–Schmidt orthogonalization). The error in the most negative exponent increases with dimension but is a numerical artefact. This figure yields strong evidence that in the region of parameter space where the network starts at a fixed point (all negative Lyapunov exponents), grows to having the maximum number of positive exponents and returns to having a few positive exponents, the variation in any specific Lyapunov exponent is very small.

There is a specific relationship between the above data and definition 12; num-Lipschitz is a stronger condition than num-continuity of Lyapunov exponents. The mean num-continuity at $n = 32$, $d = 4$

$$|\chi_j(s + \delta_{\text{num}}) - \chi_j(s)| < k\delta_{\text{num}} \tag{30}$$

$$|0.02| < k|0.01|, \tag{31}$$

yielding $k = 2$ which would not classify as num-Lipschitz contracting, whereas for $n = 32$, $d = 128$ we arrive at

$$|\chi_j(s + \delta_{\text{num}}) - \chi_j(s)| < k\delta_{\text{num}} \tag{32}$$

$$|0.004| < k|0.01|, \tag{33}$$

which yields $k = 0.4 < 1$ which does satisfy the condition for num-Lipschitz contraction. Even more striking is the num-continuity of only the largest Lyapunov exponent; for $n = 32$, $d = 4$ we get

$$|\chi_j(s + \delta_{\text{num}}) - \chi_j(s)| < k\delta_{\text{num}} \tag{34}$$

$$|0.015| < k|0.01|, \tag{35}$$

[10] The mean num-continuity for $d = 4$ and $d = 128$ is 0.015 ± 0.03 and 0.004 ± 0.003, respectively. The mean num-continuity of the largest exponent for $d = 4$ and $d = 128$ is 0.01 ± 0.03 and 0.002 ± 0.004, respectively. The discrepancy between these two data points comes from the large error in the negative exponents at high dimension.

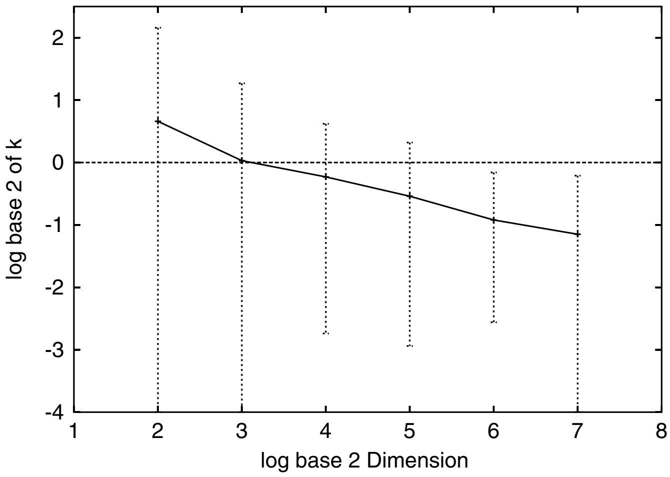

Figure 11. k-scaling: $\log_2$ of dimension versus $\log_2$ of num-Lipschitz constant of the largest Lyapunov exponent.

which yields $k = 1.5$, while the $n = 32$ $d = 128$ case is

$$|\chi_j(s + \delta_{\text{num}}) - \chi_j(s)| < k\delta_{\text{num}} \tag{36}$$

$$|0.002| < k|0.01|, \tag{37}$$

which nets $k = 0.2$. As the dimension is increased, k decreases; and thus num-continuity increases. As can be seen from figure 10, the num-continuity is achieved rather quickly as the dimension is increased; the Lyapunov exponents are quite continuous with respect to parameter variation by 16 dimensions. For an understanding of an asymptotic limit of high dimension, consider figure 11. As the dimension is increased, the $\log_2$ of the dimension versus the $\log_2(k_{\chi_1})$ yields the scaling $k \sim \sqrt{(2/d)}$; thus, as $d \to \infty$, $k_{\chi_1} \to 0$, which is exactly what we desire for continuity in the Lyapunov exponents versus parameter change.

5.1.3. Relevance. Conjectures (1), (2) and (4) are all fundamentally based on condition (1). For the neural networks, all we need to establish conjecture (1) is the num-continuity of the Lyapunov exponents, the existence of the fixed point for s near 0, the periodic orbits for $s \to \infty$ and three exponents that are, over some region of parameter space, all simultaneously positive. The n-continuity of Lyapunov exponents implies, within numerical precision, that Lyapunov exponents both pass through zero (and do not jump from positive to negative without passing through zero) and are zero within numerical precision.

5.2. *a-density of zero crossings*

Many of our arguments will revolve around varying s in a range 0.1–6 and studying the behaviour of the Lyapunov spectrum. One of the most important features of the Lyapunov spectrum we will need is a uniformity in the distribution of positive exponents between 0 and $\chi_{\max}$. As we are dealing with a countable set, we will refrain from any measure theoretic notions and instead rely on the a-density of the set of positive exponents as the dimension is increased. Recall the definition of a-dense (definition (15)), the definition of a bifurcation chain subset (definition (18)), which corresponds to the set of Lyapunov exponent zero crossings, and the definition of a chain link set (definition (16)). Our conjectures will make sense if and only if,

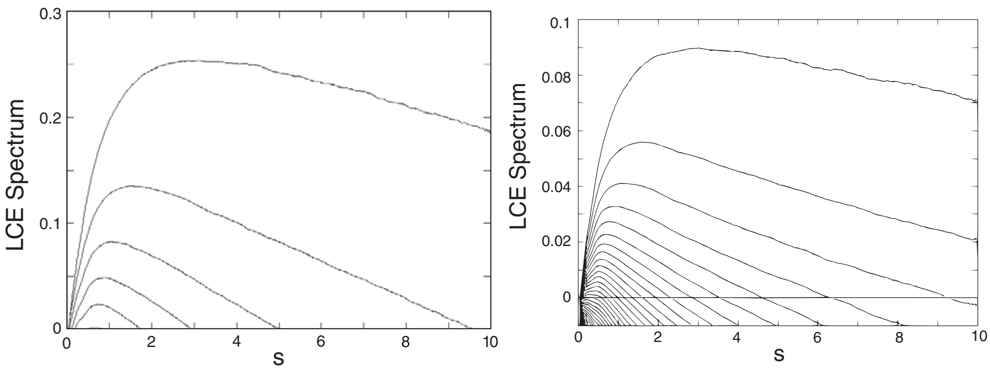

Figure 12. Positive LE spectrum for typical individual networks with 32 neurons and 16 (left) and 64 (right) dimensions.

as the dimension is increased, the bifurcation chain subsets become 'increasingly' dense or a-dense in the closure of the chain link set ($\bar{V}$). The notion of an a-dense bifurcation chain set in the closure of the chain link set as the dimension which is increased provides us with the convergence to density of non-hyperbolic points we need to satisfy our goals.

5.2.1. Qualitative analysis. To see qualitatively why we believe the a-density of Lyapunov exponent zero-crossings (a-dense bifurcation chain set in the closure of the chain link set) over a particular region of parameter space exists, consider the 16-dimensional case shown in figure 12. Begin by splitting the s variation into two regions $R_I = [0, 0.5]$, and $R_{II} = [0.5, 10]$. We then partition R_{II} using the bifurcation link sets and collect the zero crossings in the bifurcation chain sets. We want the elements of the bifurcation chain sets to be spaced evenly enough so that, as the dimension goes to infinity, variations in s on the chain link set will lead to a Lyapunov exponent zero-crossing (and a transition from V_i to $V_{i\pm 1}$)[11]. Considering region II[12], we wish for the distance along the s axis between Lyapunov exponent zero-crossings (elements of the bifurcation chain subset) to decrease as the dimension is increased. If, as the dimension is increased, the Lyapunov exponents begin to 'bunch-up' and cease to be at least somewhat uniformly distributed, the rest of our arguments will fail. For instance, in region two of the plots in figure 12, if the Lyapunov exponents were 'clumped', there will be many holes where variation of s will not imply an exponent crossing. Luckily, considering the 64-dimensional case given in figure 12, the spacing between exponent zero-crossings decreases as the dimension is increased (consider the region [0.5, 4]), and there are no point accumulations of exponents. It is also reassuring to note that even at 16 dimensions, and especially at 64 dimensions, the Lyapunov exponents are distinct and look num-continuous as previously asserted. The above figures only show two networks; if we want a more conclusive statement, we will need statistical arguments.

5.2.2. Quantitative and numerical analysis. Our analysis that specifically targets the a-density of Lyapunov exponent zero crossings focuses on an analysis of plots of the number of positive exponents versus the s parameter.

[11] Recall, the bifurcation chain sets will not exist when the zero crossings are not transverse.
[12] We will save region I for a different report. For insight into some of the dynamics and phenomena of region I, see [37].

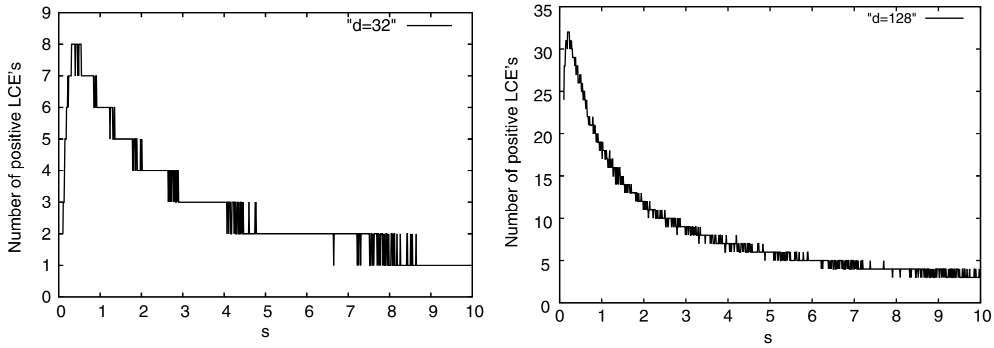

Figure 13. Number of positive LEs for typical individual networks with 32 neurons and 32 (left) and 128 (right) dimensions.

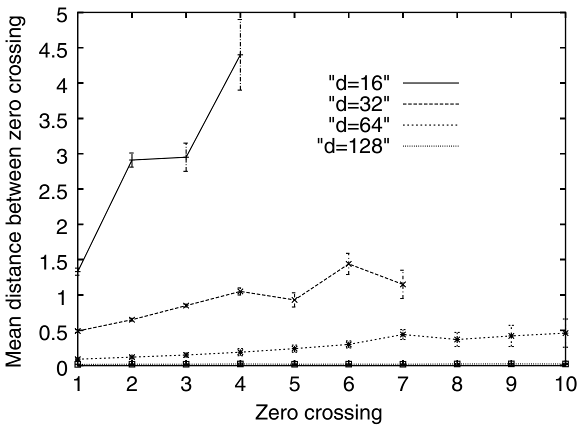

Figure 14. Mean distance between the first 10 zero crossings of LEs for many networks with 32 neurons and 16, 32, 64 and 128 dimensions.

Qualitatively, the two examples given in figure 13 (both of which typify the behaviour for their respective number of neurons and dimensions) exemplify the a-density for which we are searching. As the dimension is increased, the plot of the variation in the number of positive exponents versus s becomes more smooth[13], while the width of the peak becomes more narrow. Thus, the slope of the number of positive exponents versus s between $s = s_*$ (s_* is s where there exists the maximum number of positive Lyapunov exponents) and $s = 2$ drops from -3 at $d = 32$ to -13 at $d = 128$. Note the more negative the slope, the less the variation in s required to force a zero-crossing; this implies a-density of zero-crossings. We will not take that line of analysis further, but rather will give brute force evidence for a-density by directly noting the mean distance between exponent zero-crossings.

From figure 14, it is clear that as the dimension of the network is increased, the mean distance between successive exponent zero-crossings decreases. Note that measuring the mean distance between successive zero-crossings both in an intuitive and brute force manner verifies the sufficient condition for the a-density of the set of s values for which there exist

[13] This increase in smoothness is not necessarily a function of an increased number of exponents. A dynamical system that undergoes massive topological changes upon parameter variation will not have a smooth curve such as in figure 13, regardless of the number of exponents.

zero-crossings of exponents. The error bars represent the standard deviation of the length between zero-crossings over an ensemble (several hundred for low dimensions and on the order of a hundred for $d = 128$) of networks. For the cases where the dimension was 16 and 32, the s increment resolution was $\delta s = 0.01$. The error in the zero-crossing distance for these cases is, at the smallest, 0.02, while at its smallest, the zero-crossing distance is 0.49; thus, resolution of 0.01 in s is sufficient to adequately resolve the zero-crossings. Such is not the case for 64- and 128-dimensional networks. For these cases we were required to increase the s resolution to 0.005. The zero-crossings of a few hundred networks considered were all examined by hand; the distances between the zero-crossings were always distinct, with a resolution well below that necessary to determine the zero-crossing point. The errors were also determined by hand, noting the greatest and least points for the zero-crossing. All the zero-crossings were determined after the smallest positive exponent, which became positive, hit its peak value, i.e. after approximately 0.75 in the $d = 16$ case of figure 12.

5.2.3. Relevance. The a-density of zero-crossings of Lyapunov exponents provides the most important element in our arguments of conjectures (1) and (2); combining num-continuity with a-density will essentially give our desired results. If continuity of Lyapunov exponents increases, and the density of zero crossings of exponents increases over a set $U \in R^1$ of parameter space, it becomes clear that we will have both hyperbolicity violation and, upon variation of parameters in U, we will have the topological change we are claiming. Of course, small issues remain, but these will be dealt with in the final arguments.

5.3. Arbitrarily large number of positive exponents

For our a-density arguments to work, we need a set whose cardinality is asymptotically a countably infinite set (such that it can be a-dense in itself) and we need the distance between the elements in the set to approach zero. The latter characteristic was the subject of the previous section; the former subject is what we intend to address in this section.

5.3.1. Qualitative analysis. The qualitative analysis of this can be seen in figure 13; as the dimension is increased, the maximum number of positive Lyapunov exponents clearly increases.

5.3.2. Quantitative analysis. Figure 15 depicts the number of positive Lyapunov exponents versus dimension, from which it is clear that as the dimension is increased, the number of positive exponents increases in a nearly linear fashion [14]. Further, this plot is linear to as high a dimension as we could compute enough cases for reasonable statistics. This scaling is dependent on the number of neurons in that an increase in N increases the slope (for more information see [34]).

5.3.3. Relevance. The importance of the increasing number of positive exponents with dimension is quite simple. For the a-density of exponent zero crossing to be meaningful in the infinite-dimensional limit, there must also be an arbitrarily large number of positive exponents that can cross zero. If, asymptotically, there is a finite number of positive exponents, all of our claims will be false; a-density requires a countably infinite set.

[14] Further evidence for such an increase is provided by considering the Kaplan–Yorke dimension versus d. Such analysis yields a linear dependence, $D_{K-Y} \sim d/2$.

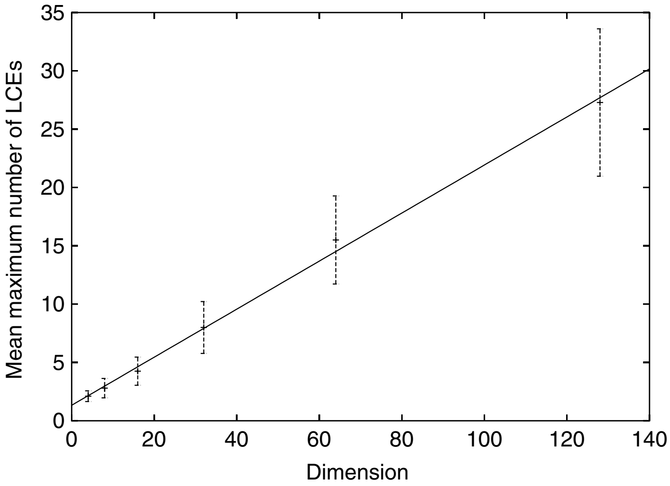

Figure 15. Mean maximum number of positive LEs versus dimension, all networks have 32 neurons (slope is approximately $\frac{1}{4}$).

6. Numerical arguments for conjectures

6.1. Decreasing window probability

With the num-continuity and a-density arguments already in place, all the evidence required to show the length of periodic windows along a curve in parameter space is already in place. We will present some new data, but primarily we will clarify exactly what the conjecture says. We will also list the specifics under which the conjecture applies in our circumstances.

6.1.1. Qualitative analysis. Qualitative evidence for the disappearance of periodic windows amidst chaos is evident from figures 7, 8 and 12; the periodic windows that dominate the 4-dimensional network over the parameter range $s = 0$–10 are totally absent in the 64-dimensional network. It is important to note that for this conjecture, as well as all our conjectures, we are considering the s parameter over ranges no larger than 0–10. We will avoid, for the most part, the 'route to the chaos' region (s near zero), as it yields many complex issues that will be saved for another paper [36, 38]. We will instead consider the parameter region after the lowest positive exponent first becomes positive. We could consider parameter ranges considerably larger, but for s very large, the round-off error begins to play a significant role, and the networks become binary. This region has been briefly explored in [37]; further analysis is necessary for a more complete understanding [91].

6.1.2. Quantitative and numerical analysis. The quantitative analysis will involve arguments of two types; those that are derived from data given in sections 5.1 and 5.2 and those that follow from statistical data regarding the probability of a window existing for a given s along an interval in R.

The conjecture we are investigating claims that as the dimension of a dynamical system is increased, periodic windows along a one-dimensional curve in parameter space vanish in a significant portion of parameter space for which the dynamical system is chaotic. This is, of course, highly dependent upon the region of parameter space one is observing. For our purposes, we will generally be investigating the region of s parameter space between 0.1 and 10; however, sometimes we will limit the investigation to s between 2 and 4. Little changes if we increase s until the network begins behaving as a binary system due (quite possibly) to the round-off error. However, along the transition to the binary region, there are significant complications which we will not address here. As the dimension is increased, the main concern is that the lengths of the bifurcation chain sets must increase such that there will exist at least one bifurcation chain set that has a cardinality approaching infinity as the dimension of the network approaches infinity.

Our first argument is based directly upon the evidence of num-continuity of Lyapunov exponents. From figure 10 it is clear that as the dimension of the set of networks sampled is increased, the mean difference in Lyapunov exponents over small ($\delta s = 0.01$) s parameter perturbation decreases. This increase in num-continuity of the Lyapunov exponents with dimension over the parameter range is a direct result of the disappearance of periodic windows from the chaotic regions of parameter space. This evidence is amplified by the decrease in the standard deviation of the num-continuity versus dimension (of both the mean of the exponents and the largest exponent). This decrease in the standard deviation of the num-continuity of the largest Lyapunov exponent allows for the existence of fewer large deviations in Lyapunov exponents (large deviations are needed for all the exponents to suddenly become less than or equal to zero).

We can take this analysis a step further and simply calculate the probability of an s value having a periodic orbit over a given interval. Figure 16 shows the probability of a periodic window existing for a given s on the interval $(2, 4)$ with $\delta s = 0.001$ for various dimensions. There is a power law in the probability of periodic windows—the probability of the existence of a periodic window decreases approximately as $1/d$. Moreover, in high-dimensional dynamical systems, when periodic windows are observed on the interval $(2, 4)$, they are usually large. In other words, even though the probability that a given s value will yield a periodic orbit for $d = 64$ is 0.02, it is likely that the probability is contained in a single connected window, as opposed to the lower dimensional scenario where the probability of window occurrence is distributed over many windows. We will save further analysis of this conjecture for a different report ([65]), but hints as to why this phenomenon is occurring can be found in [92].

6.1.3. Relevance. Decreasing window probability inside the chaotic region provides direct evidence for conjecture (3) along a one-dimensional interval in parameter space. We will use the decreasing periodic window probability to help verify conjecture (2) since it provides the context we desire with the num-continuity of the Lyapunov spectrum. Our argument requires that there exists at least one maximum in the number of positive Lyapunov exponents with parameter variation. Further, that maximum must increase monotonically with the dimension of the system. The existence of periodic windows causes the following problems: periodic windows can still yield structural instability—but in a catastrophic way; periodic windows split up the bifurcation chain sets which, despite not being terminal to our arguments, provide many complications with which we do not contend. However, we do observe a decrease in periodic windows, and with the decrease in the (numerical) existence of periodic windows comes the decrease in the number of bifurcation chain sets; i.e. $l = |a_n - a_1|$ is increasing yet will remain finite.

6.2. Hyperbolocity violation

We will present two arguments for hyperbolicity violation—or nearness to hyperbolicity violation of a map at a particular parameter value, s. The first argument will consider the

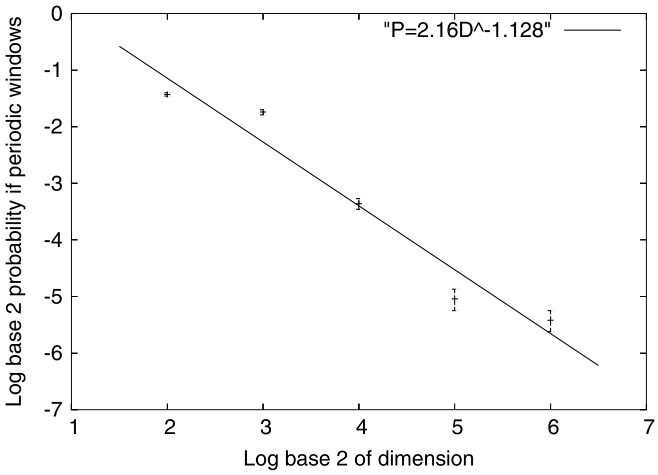

Figure 16. $\log_2$ of the probability of periodic or quasi-periodic windows versus $\log_2$ of dimension. The line $P_w = 2.16d^{-1.128}$ is the least squares fit of the plotted data.

fraction of Lyapunov exponents near zero over an ensemble of networks versus variation in s. If there is any hope of the existence of a chain link set with bifurcation link sets of decreasing length, our networks (on the s interval in question) must always have a Lyapunov exponent near zero. The second argument will come implicitly from a-density arguments presented in section 5.2. To argue for this conjecture, we only need the existence of a neutral direction[15], or, more accurately, at least two bifurcation link sets, which is not beyond reach.

6.2.1. Qualitative analysis. A qualitative analysis of hyperbolocity violation comes from combining the num-continuity of the exponents in figure 8 and the evidence of exponent zero crossings from figures 13 and 10. If the exponents are continuous with respect to parameter variation (at least locally) and they start negative, become positive and eventually become negative, then they must be zero (within numerical precision) for at least two points in the parameter space. It happens that the bifurcation chain link sets are LCE decreasing from i to $i + 1$, which will provide additional, helpful structure.

6.2.2. Quantitative and numerical analysis. The first argument, which is more of a necessary but not sufficient condition for the existence of hyperbolicity violation, consists of searching for the existence of Lyapunov exponents that are zero within the allowed numerical errors. With num-continuity, this establishes the existence of exponents that are numerically zero. For an intuitive feel for what numerically zero means, consider the oscillations in figure 13 of the number of positive exponents versus parameter variation. It is clear that as they cross zero there are numerical errors that cause an apparent oscillation in the exponent[16]. There is a certain fuzziness in numerical results that is impossible to remove. Thus, questions regarding exponents being exactly zero are ill-formed. Numerical results of the type presented in this paper need to be viewed in a framework similar to physical experimental results. With this in mind, we note the significance of the exponents near zero. To do this, we calculate the relative

[15] By neutral direction we mean a zero Lyapunov exponent; we do not wish to imply that there will always exist a centre manifold corresponding to the zero Lyapunov exponent.

[16] It is possible that there exist Milnor style attractors for our high-dimensional networks or at least multiple basins of attraction. As this issue seems to not contribute, we will save this discussion for a different report.

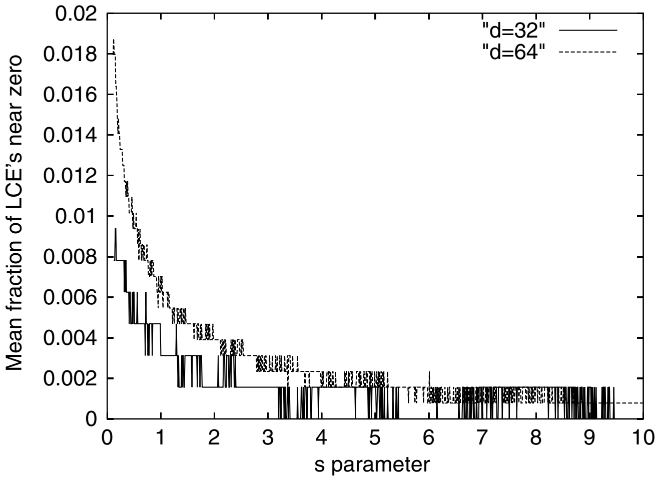

Figure 17. Mean fraction of LEs near zero (0 ± 0.01) for networks with 32 neurons and 32 or 64 dimensions (averaged over 100 networks).

number of Lyapunov exponents numerically at zero compared with the ones away from zero. All this information can be summarized in figure 17, which addresses the mean fraction of exponents that are near zero.

The cut-off for an exponent being near zero is ± 0.01, which is approximately the expected numerical error in the exponents for the number of iterations used. There are four important features to notice about figure 17: (i) there are no sharp discontinuities in the curves; (ii) there exists an interval in parameter space such that there is always at least one Lyapunov exponent in the interval $(-0.01, 0.01)$, and the length of that parameter interval is increasing with dimension; (iii) the curves are concave—implying that exponents are somehow leaving the interval $(-0.01, 0.01)$ and (iv) there is a higher fraction of exponents near zero at the same s value for higher dimension. The first property is important because holes in the parameter space where there are no exponents near zero would imply the absence of the continuous zero crossings we will need to satisfy conjecture (2). To satisfy conjecture (1) we only need three exponents to be near zero and undergo a zero crossing for the minimal bifurcation chain subset[17] to exist. There are clearly enough exponents on average for such to exist for at least some interval in parameter space at $d = 32$, e.g. for $(0.1, 0.5)$. For $d = 64$ that interval is much longer—$(0.1, 1)$. Finally, if we want the chain link set to be more connected and for the distance between elements of the bifurcation chain subset to decrease, we will need the fraction of exponents near zero for the fixed interval $(-0.01, 0.01)$ for a given interval in s to increase with dimension. This figure does not imply that there will exist zero-crossings, but it provides the necessary circumstance for our arguments.

The second argument falls out of the a-density and num-continuity arguments. We know that as the dimension is increased, the variation of Lyapunov exponents versus parameter variation decreases until, at dimension 64, the exponent varies continuously within numerical errors (and thus upon moving through zero, the exponent moves through zero continuously). We also know that on the interval in parameter space $A = [0.1, 6]$, the distance between exponent zero crossings decreases monotonically. Further, on this subset A, there always exists a positive Lyapunov exponent, thus implying the existence of a bifurcation chain set whose length is at least 5.9. Extrapolating these results to their limits in infinite dimensions,

[17] The minimal bifurcation chain subset requires at least two adjoining bifurcation link sets to exist.

the number of exponent crossings on the interval A will monotonically increase with dimension. As can be seen from figure 14, the exponent zero-crossings are relatively uniform with the distance between crossings decreasing with increasing dimension. Considering figure 12, the exponent zero crossings are also transverse to the s axis. Thus, the zero-crossings on the interval A, which are exactly the points of non-hyperbolocity we are searching for, become dense. This is overkill for the verification of the existence of a minimal bifurcation chain set. This is strong evidence for both conjectures (1) and (2). It is worth noting that hitting these points of hyperbolocity violation upon parameter variation is extremely unlikely under any uniform measure on R as they are a countable collection of points.[18] Luckily, this does not matter for either the conjecture at hand or for any of our other arguments.

6.2.3. Relevance. The above argument provides direct numerical evidence of hyperbolocity violation over a range of the parameter space. This is a strong evidence supporting conjecture (1). It does not yet verify conjecture (2), but it sets the stage because we have shown that there is a significant range over which hyperbolocity is violated. The former statement speaks of conjecture (4) also; a full explanation of conjecture (4) requires further analysis, which is the subject of a discussion in the final remarks.

6.3. Hyperbolocity violation versus parameter variation

We are finally in a position to consider the final arguments for conjecture (2). To complete this analysis, we need the following pieces of information:

(i) the maximum number of positive exponents to go to infinity,
(ii) a region of parameter space for which a-density of Lyapunov exponent zero crossings exists; i.e. we need an arbitrarily large number of adjoining bifurcation link sets (such that the cardinality of the bifurcation chain set becomes arbitrarily high) such that for each V_i, the length of V_i, $l = |b_i - a_i|$, approaches zero,
(iii) num-continuity of exponents to increase as the dimension increases.

The a-density, num-continuity and the arbitrary numbers of positive exponent arguments we need have, for the most part, been provided in previous sections. In this section we will simply apply the a-density and num-continuity results in a manner that suits our needs. The evidence for the existence of a single maximum in the number of positive exponents, a mere convenience for our presentation, is evident from section 5.3. We will simply rely on all our previous figures and the empirical observation that, as the dimension is increased above $d = 32$ for networks that have the typical num-continuity (which includes all networks observed for $d \geqslant 64$), there exists a single, global maximum in the number of positive exponents versus parameter variation.

6.3.1. Qualitative analysis. The qualitative picture we are using for intution is that of figure 12. This figure displays all the information we wish to quantify for many networks; as the dimension is increased, there is a region of parameter space where the parameter variation needed to achieve a topologically different (by topologically different, we mean a different number of global stable and unstable manifolds) attractor decreases to zero. Based on figure 12 (and hundreds of similar plots), we claim that this parameter range exists for at least $0.5 \leqslant s \leqslant 6$.

[18] When considering parameter values greater than the point where the smallest exponent becomes positive, the zero crossings seem always to be transverse. For smaller parameter values—along the route to chaos, a much more complicated scenario ensues.

6.3.2. Quantitative and numerical analysis. Let us now complete our arguments for conjecture (2). For this we need a subset of the parameter space, $B \subset R^1$, such that some variation of $s \in B$ will lead to a topological change in the map f in the form of a change in the number of global stable and unstable manifolds. Specifically, we need $B = \bigcup V_i = V$, where V_i and V_{i+1} share a limit point and are disjoint. Further, we need the variation in s needed for the topological change to decrease monotonically with dimension on V. More precisely, on the bifurcation chain set, U, the distance between elements must decrease monotonically with increasing dimension. We will argue in three steps: first, we will argue that, for each f with a sufficiently large number of dimensions, there will exist an arbitrarily large number of exponent zero crossings (equivalent to an arbitrarily large number of positive exponents); next we will argue that the zero crossings are relatively smooth and finally, we will argue that the zero crossings form an a-dense set on V—or on the bifurcation chain set, $l = |b_i - a_i| \to 0$ as $d \to \infty$. This provides strong evidence supporting conjecture (2).

Assuming a sufficiently large number of dimensions, verification of conjecture (1) gives the existence of the bifurcation chain set and the existence of the adjoining bifurcation link sets. The existence of an arbitary number of positive Lyapunov exponents and thus an arbitrarily large number of zero-crossings follows from section 5.3. That the bifurcation chain set has an arbitrarily large number of elements, $\#U \to \infty$ is established by conjecture (3) because, without periodic windows, every bifurcation link set will share a limit point with another bifurcation link set. From section 5.1, the num-continuity of the exponents persists for a sufficiently large number of dimensions; thus, the Lyapunov exponents will cross through zero. Finally, section 5.2 tells us that the Lyapunov exponent zero-crossings are a-dense; thus, for all $c_i \in U$, $|c_i - c_{i+1}| \to 0$, where c_i and c_{i+1} are sequential elements of U. For our work, we can identify U as $U \subset [0.5, 6]$. We could easily extend the upper bound to much greater than 6 for large dimensions ($d \geqslant 128$). How high the upper bound can be extended will be discussed in further work. It is useful to note that the bifurcation link sets are LCE decreasing with increasing s. This is not necessary to our arguments, but it is a nice added structure that aids our intuition. The LCE decreasing property exists due to the existence of the single, global maximum in the maximum number of positive Lyapunov exponents followed by an apparent exponential decrease in the number of positive Lyapunov exponents.

6.3.3. Relevance. The above arguments provide direct evidence of conjectures (2) and (4) for a one-dimensional curve (specifically an interval) in parameter space for our networks. This evidence also suggests the persistence of chaos in high-dimensional networks with perturbations on higher-dimensional surfaces in parameter space. Finally, despite the seemingly inevitable topological change upon minor parameter variation, the topological change is quite benign.

7. Fitting everything together

Having finished with our specific analysis, we now put our work in the context of other work, both of a more mathematical and a more practical and experimental nature. In this spirit, we will provide a brief summary of our arguments followed by a discussion of how our results mesh with various theoretical results from dynamical systems and turbulence.

7.1. Summary of arguments

7.1.1. Periodic window probability decreasing: conjecture 3. The conjecture that the probability of periodic windows for a given s value along an interval in parameter space decreases with increasing dimension upon the smallest positive Lyapunov exponent becoming positive is initially clear from considering the Lyapunov spectra of neural networks versus parameter variation for networks of increasing size (figures 7 and 8). We show that as the dimension is increased, the observed probability of periodic windows decreases inversely with increasing dimension. This analysis is independent of the num-continuity analysis, and the results from the analysis of num-continuity and periodic window probability decrease reinforce each other. The mechanism that this conjecture provides is the lengthening of the bifurcation chain set.

Further investigations of this phenomenon will follow in a later report. For other related results see [35, 65, 92–94].

7.1.2. Hyperbolicity violation: conjecture 1. The intuition for this conjecture arises from observing that for high-dimensional systems, there exists at least one Lyapunov exponent that starts negative, becomes positive, then becomes negative again; thus, if it behaves numerically continuously, it must pass through zero for some parameter value s.

To verify this conjecture, we presented two different arguments. The first argument was a necessary but not sufficient condition for hyperbolicity violation. We show that over a sizeable interval in parameter space there exists a Lyapunov exponent very near zero, and the fraction of the total number of Lyapunov exponents that are near zero increases over a larger interval of parameter space as the dimension is increased. The second argument was based on the a-density of exponent zero crossings, the num-continuity of the exponents as the dimension increased and the increasing number of positive exponents with dimension. Both arguments together help imply an interval of parameter space such that on that interval the number of parameter values such that hyperbolicity is violated is increasing.

It is important to note that it is possible that hyperbolicity can be violated along the orbit. If hyperbolicity is violated along the orbit for a measurable set of values, but the mean of each of the exponents averaged along the orbit is non-zero, then the diagnostics in this paper will be inconclusive. Insight into hyperbolicity violation along a given orbit can be obtained by considering a histogram of the exponents and observing whether the tails of the distributions of the exponents near zero overlap zero. Such an analysis is done on a case by case basis. Because this paper represents a statistical study, we have reserved such an analysis for another paper.

7.1.3. Existence of codimension-ϵ bifurcation set: conjecture 2. The intuition for this argument follows from observing that the peak in the number of positive Lyapunov exponents tends toward a spike of increasing height and decreasing width as the dimension is increased. This, with some sort of continuity of exponents, argues for a decrease in distance between exponent zero crossings.

Summarizing the arguments regarding conjecture (2), with increasing dimension we have increased num-continuity of Lyapunov exponents; increasing number of positive Lyapunov exponents and a-density of Lyapunov exponent zero crossings (thus all the exponents are not clustered on top of each other). Thus, on a finite set in parameter space, there is an arbitrary number of exponents that move smoothly from negative values to positive values and back to negative values. Further, these exponents are relatively evenly spaced. Thus, the set in parameter space for which hyperbolicity is violated is increasingly dense; and with an arbitrary number of violations available, the perturbation of the parameter required to force a topological change (a change in the number of positive exponents) becomes small.

Further evidence of continuous topological change is provided by the smooth variation of the plot of the Kaplan–Yorke dimension versus s that corresponds to figure 13 [35].

7.1.4. Non-genericity of structural stability: conjecture 4. As previously mentioned, it could appear that our results contradict Robbin [3], Robinson [4] and Mañé [5]. We will discuss specifically how our results fit with theirs in section 7.3.1. In the current discussion, we interpret our results in a numerical context.

We claim to have found a subset of parameter space that, in the limit of infinite dimensions, has dense hyperbolicity violation. This could be interpreted to imply that we have located a set for which strict hyperbolicity does not imply structural stability because the C^1 changes in the parameter give rise to topologically different behaviour. The key issue is that in numerical simulations, there do not exist infinitesimals or infinite-dimensional limits[19]. Rather, we can speak as to how behaviour arises and how limits behave along the path to the ideal. We have found a subset of parameter space that can approximate (with unlimited computing) arbitrarily closely a set for which hyperbolicity will not imply structural stability. Thus, an experimentalist or a numerical scientist might see behaviour that apparently violates the results of Robbin [3], Robinson [4] and Mañé [5]; yet it will not be strictly violating those theorems. The key point of this conjecture is that we can *observe* apparent violation of the structural stability conjecture, but the violation (on a Lebesgue measure zero set) occurs as smooth, not catastrophic, topological change. (In section 7.3.1 we will further discuss our results as they relate to those of Robbin [3], Robinson [4] and Mañé [5].)

7.2. Fitting our results in with the space of C^r functions: how our network selection method affects our view of function space

The act of performing a numerical experiment induces a measure on whatever is being experimented upon. A measure, in a very general sense, provides a method of determining the volume that a set occupies in its ambient space (for a formal treatment, see [95]). Usually that method provides a specific mechanism of measuring lengths of a covering interval. Then, the entire space is covered with the aforementioned intervals, and their collective volume is summed. One of the key issues is how the intervals are weighted. For instance, considering the real line with the standard Gaussian measure imposed upon it; the interval $[-1, 1]$ contains the majority of the volume of the entire interval $[-\infty, \infty]$. Our method of weighting networks selects fully connected networks with random Gaussian weights. Thus, in the limit of high dimension and large number of neurons, very weakly connected networks will be rare because the Gaussian statistics of the weights will be dominant. Likewise, fully connected networks where all the weights have the same strength (up to an order of magnitude) will also be uncommon. One can argue whether or not the measure we utilize realistically represents the function space of nature, but those arguments are fundamentally ill-formed because they cannot be answered without either specific information about the natural system with which our framework is being compared or the existence of some type of invariant measure. The best hope for specifying connections between natural systems, other computational models and a neural network construction such as the point presented here is using the aforementioned notion of prevalence. Nevertheless, the product measure we impose does cover the entire space of neural networks noted in section 1, although all sets do not have equal likelihood of being selected, and thus our results must be interpreted with this in mind.

A second key issue regards how the ambient space is split into intervals or, in a numerical sense, how the grain of the space is constructed. We will again introduce a simpler case for purposes of illustration, followed by a justification of why the simpler case and our network framework are essentially equivalent. Begin with R^n and select each coordinate (v_i) in the

[19] Note that in previous sections we do use words like, 'in the infinite-dimensional limit'. It is for reasons such as these (and many others) that we are only putting forth conjectures and not theorems; this distinction is not trivial.

vector $\boldsymbol{v} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\} \in \boldsymbol{R}^n$ from a normal, i.i.d. distribution with mean zero, variance one. Next, suppose that we are attempting to see every number and every number combination. This will be partially achieved by the random number selection process mentioned above, and it is further explored by sweeping the variance, i.e. selecting a scalar $s \in R$, $0 \leqslant s$ and sweeping s over the positive real line, $s\boldsymbol{v}$. This establishes two meshes, one for the individual vectors, which is controlled by how finely the s parameter is varied, and another mesh that controls how the initial coordinates are selected. These two combined meshes determine the set of combinations of coordinates that will be observed. If one considers how this affects vector selection in, say, R^3, for simplicity, both in the initial vector selection and in the vector sweeping, it is clear how R^3 will be carved out. The point is that we can directly associate how we carve up our neural network function space with how we carve up the neural network weight space. To understand how our neural network selection process works, simply associate $\boldsymbol{v}$ with the vectors in the ω matrix and s with the scaling parameter s. This keeps the view of our function space largely in standard Euclidean space. Of course there is the last remaining issue of the amplitude terms, the β. Apply the same type of analysis to the β as for the ω. Of course, initially it would seem that the scaling parameter is missing, but note that multiplying the β by s, in our networks, is essentially equivalent to multiplying the ω by s. To understand this, consider the one-dimensional network, with one neuron:

$$x_t = \beta_0 + \beta_1 \tanh(s\omega_0 + s\omega_1(\beta_0 + \beta_1 \tanh(s\omega_0 + s\omega_1 x_{t-1}))). \tag{38}$$

It is clear from this that inserting s inside tanh will sweep the β, but inserting s outside the squashing function will miss sweeping the ω_0 bias term.[20] What this all amounts to is a heuristic construction of a probe space (relative to the measures on the parameter space) for embeddings of C^r dynamical systems using neural networks. More formal constructions are topics of current work.

Summarizing, it should be clear that the measures we utilize will capture the entire space of neural networks we are employing. Yet, it should also be clear that we will not select each network with equal probability. Moreover, the measure we impose is a product measure, not a joint probability measure where there is a co-dependence between the parameters. This hopefully has a significant impact on our conclusions and results—for there is too much diversity in natural systems to be caught by a single measure on a space of universal approximators. A complete connection between network structure and dynamics, in an understandable language, is yet out of reach (as opposed to, say, for spherical harmonics), but this is a current direction of research. Various research programs for quantifying various spaces of neural networks and their dynamics by the measures on the weights are currently underway using tools such as information geometry [53, 96]. Nevertheless, we claim that our framework gives a complete picture of the space of C^r maps of compact sets to compact sets with the Sobolev metric from the perspective of a *particular* product measure.

7.3. Our results related to other results in dynamical systems

As promised throughout, we will now connect our results with various theorems and conjectures in the field of dynamical systems. This will hopefully help put our work in context and increase its understandability.

[20] Recalling section 2.1.4, we actually do a little more with the β than we mention here; the previous argument is simply meant to give a (mathematically imprecise) picture of how our experiment carves out the space of neural networks.

7.3.1. Structural stability theory and conjecture 4. It is now time to address the apparent conflict between our observations and the structural stability theorems of Robbin [3], Robinson [4] and Mañé [5]. We would like to begin by noting that we do not doubt the validity or correctness of any of these results. In fact, any attempt to use our techniques and results to provide a counter example to the theorems of Robbin, Robinson or Mañé involves a misunderstanding of what our methods do and imply.

In conjecture (4) we claim, in an intuitive sense, that along a one-dimensional curve in parameter space, our dynamical systems are hyperbolic with Lebesgue measure one. Yet, we can still find subsets that are measure zero, yet a-dense, for which our dynamical systems are partially hyperbolic rather than hyperbolic. The motivation for the above statement roughly derives from thinking of a turbulent fluid. In this circumstance, the number of unstable manifolds can be countably infinite, and upon varying, say, the viscosity, from very low to very high, one would have a countable number of exponents becoming positive over a finite length of parameter space. Yet, all the limits of this sort and all the intuition about what will happen in the infinite-dimensional limit are just ideas. There are limits to what we can compute; there do not exist infinite-dimensional limits or infinitesimals in numerical computing, and aside from the existence of convergence theorems, we are unable to draw conclusions beyond what the data say. Thus, our results do not provide a counter-example to the stability conjecture. Rather, a key point of our results is that we do observe, in a realistic numerical setting, structural instability upon small parameter variation. It is useful to think instead of structural stability as an open condition on our parameter space whose endpoints correspond to the points of structural instability—the points of bifurcations in turbulence. These disjoint open sets are precisely the bifurcation link subsets, V_i, for which the map f is structurally stable. As the dimension is increased, the lengths of the V_i decrease dramatically and may fall below numerical or experimental resolution. Thus, the numerical or experimental scientist might observe topological variation in the form of a variation in the number of positive Lyapunov exponents upon parameter variation in systems that should be (and technically are) structurally stable according to the work of Robbin, Robinson and Mañé; i.e. the scientist might observe structural instability. This is the very practical difference between numerical computing and the world of strict mathematics. (Recall we were going to connect structural stability theory closer to reality; the former statement is as far as we will go in this paper.) The good news is that even though observed structural stability might be lost, it is lost in a very meek manner—the topological changes are very slight, just as seems to be observed in many turbulent experimental systems. Further, partial hyperbolicity is not lost, and the dynamically stable characteristics of stable ergodicity seem to be preserved; although, we obviously cannot make a strict mathematical statement.

Thus, rather than claiming our results are contrary to those of Robbin [3], Robinson [4], and Mañé [5], our results speak to both what might be seen of those theorems in high-dimensional dynamical systems and how their results are approached upon increasing the dimension of a dynamical system.

It is worth noting that, given a typical 64-dimensional network, if we fixed s at such a point that there was an exponent zero crossing, we believe (based on preliminary results) that there will exist many perturbations of other parameters that leave the exponent zero crossing unaffected. However, these perturbations are apparently of very small Lebesgue measure and of a small codimension set in parameter space, i.e. we believe we can find, in the construction we use, perturbations that will leave the seemingly transversal intersection of an exponent with 0 at a particular s value unchanged; yet, these parameter changes must be small. A very interesting question is whether the same could be said if the measure on the weight space was a joint rather than a product measure.

7.3.2. Stable ergodicity. In this study we are particularly concerned with the interplay, along a parametrized curve, of how often partial hyperbolicity is encountered versus strict hyperbolicity. It should be noted that if a dynamical system is hyperbolic, it is partially hyperbolic. All the neural networks we considered were at least partially hyperbolic; we found no exceptions. Many of the important questions regarding partially hyperbolic dynamical systems lie in showing the conditions under which such systems are stably ergodic. We will now discuss this in relation to our results and methods.

Pugh and Shub [60] put forth the following conjecture regarding partial hyperbolicity and stable ergodicity.

Conjecture 5 (Pugh and Shub [60] conjecture 3). *Let $f \in \mathrm{Diff}^2_\mu(M)$ where M is compact. If f is partially hyperbolic and essentially accessible then f is ergodic.*

The strongest result that has been shown to date regarding their conjecture is in [97].

Theorem 3 (Stable ergodicity theorem (theorem 0.1 [97])). *Let f be C^2, volume preserving, partially hyperbolic and centre bunched. If f is essentially accessible then f is ergodic and, in fact, has the Kolmogorov property.*

A diffeomorphism is partially hyperbolic if it satisfies the conditions of definition (7). Ergodic behaviour implies that, upon breaking the attractor into measurable sets, A_i, for f applied to each measurable set for enough time, $f^n(A_i)$ will intersect every other measurable set A_j. This implies a weak sense of recurrence; for instance, quasi-periodic orbits, chaotic orbits, and some random processes are at least colloquially ergodic. More formally, a dynamical system is ergodic if and only if almost every point of each set visits every set with positive measure. The accessibility property simply formalizes a notion of one point being able to reach another point. Given a partially hyperbolic dynamical system, $f : X \to X$ such that there is a splitting on the tangent bundle $TM = E^u \oplus E^c \oplus E^s$, and $x, y \in X$, y is accessible from x if there is a C^1 path from x to y whose tangent vector lies in $E^u \cap E^s$ and vanishes finitely many times (these paths are denoted u-s-paths). The diffeomorphism f is centre bunched if the spectra of Tf (as defined in section 2.2) corresponding to the stable ($T^s f$), unstable ($T^u f$) and ($T^c f$) central directions lie in thin, well separated annuli (see [97] for the most recent and general formulation; the radii of the annuli are technical and are determined by the Hölder continuity of the diffeomorphism). Lastly, let us note that a dynamical system is called stably ergodic if, given $f \in \mathrm{Diff}^2_\mu(M)$ (again M compact), there is a neighbourhood, $f \in Y \subset \mathrm{Diff}^2_\mu(M)$, such that every $g \in Y$ is ergodic with respect to μ.

The current proofs of the Pugh–Shub stable ergodicity theorem require the dynamical system to be Lebesgue measure preserving—or non-dissipative. The primary technique of the proofs is to very carefully select a positive measure set and then very carefully move it along the u-s-paths in such a way that all the positive measure sets are visited while the original set retains positive measure. Doing this hinges on absolute continuity with respect to Lebesgue measure on the orbit with respect to *both* the stable and unstable manifolds. Needless to say, it is difficult to prove that a dissipative dynamical system is ergodic. But, while the *arguments* hinge on absolute continuity on both the stable and unstable manifolds, it is believed that the *results* do not. The results from this paper reinforce this viewpoint. One tactic of weakening the area-preserving condition is to consider partially hyperbolic dynamical systems with negative (or positive) central exponents *along* the orbit (e.g. modified Pesin theory). In this vein, using theorems 10 and 11 and corollary 12 of [79], it can be concluded that partially hyperbolic dynamical systems with negative central exponents and with a particular SRB-like measure (see [79] for details) will be ergodic.

Numerical verification of ergodicity can be somewhat difficult because the modeller would have to watch each point and verify that eventually the trajectory returns very close to every other point on the orbit (i.e. it satisfies the Birkoff hypothesis). Doing this for a few points is, of course, possible; but doing it for a high-dimensional attractor for any sizable number of points can be extremely time consuming. For d high enough, we observe a single SRB measure. This can be seen in figure 8 where every s value has a different initial condition. If this is a robust result, and using results from Burns *et al* [79], if the central exponents can be shown to be negative, ergodicity will follow. Because we have not studied the exponents *along* the orbit in this work, we cannot use this argument to claim that ergodicity is existent and stable in the ensemble of systems we studied. Checking the accessibility criterion seems to pose similar problems; in fact, it is hoped that accessibility is the sufficient recurrence condition for ergodic behaviour. Thus, it should be no surprise that accessibility would be difficult to check numerically (it has been shown to be C^1 dense [98]). However, in essence, ergodicity is a notion capturing indivisibility of the attractor relative to a measure. The key point is that all positive measure sets are visited by all orbits with probability one. The most obvious, and it is thought, *the* primary means of splitting up an attractor is the existence of neutral directions or zero Lyapunov exponents (see the 'stacked Anosov' in [7] for a simple example). Considering figures 8, 9 and 12, and the overall argument presented, it is clear that the persistence of zero exponents with respect to parameter change in the systems we study is unlikely. Moreover, figures 4–6 demonstrate that quantities dependent upon the SRB measures for calculation converge in the uniform manner expected of an ergodic system. Of course, showing that one has explored all the variations inside the neighbourhood $(f \in) Y \subset \mathrm{Diff}^2_\mu(M)$ is impossible; thus, claiming that we have, in a mathematically rigorous way, observed stable ergodicity as the predominant characteristic would be premature. However, generalizing our results using notions of prevalence is a topic of current work and does not seem out of reach. What we can say is that we have never observed a dynamical system, within our construction, that is not on a compact set and is not partially hyperbolic; moreover, systems seemed to become more stably ergodic as the dimension increased. Thus, our results provide evidence that the conjecture of Pugh and Shub, which does not require area preservation or absolute continuity with respect to the stable and unstable manifolds, is on track. For more information with respect to the mathematics discussed above, see [7, 19, 59].

Comparing conjecture (5) with theorem (3), the required extra hypothesis for the proof of the theorem is only centre bunching of the spectrum of Tf. It is likely that the centre bunching hypothesis has been weakened as far is it can be for the current arguments. As it stands, the spectra that we observe, which are strongly reminiscent of what might be expected in turbulent-like dynamics, do appear to satisfy the current centre bunching criterion.[21] We cannot comment on its necessity. Moreover, considering similar numerical work done by others, and vaguely combining the circular law ([22–24]) with matrix product results ([99]), it is likely that the centre bunching criterion as it stands is not crucial for dissipative systems.

7.3.3. Palis's conjectures. Palis [9] stated many stability conjectures based upon the last thirty years of developments in dynamical systems that we wish to relate to the results in the present paper [9].

Conjecture 6 (Global conjecture on the finitude of attractors and their metric stability [9]).

(I) Denseness of finitude of attractors—there is a C^r ($r \geqslant 1$) dense set D of dynamics such that any element of D has finitely many attractors whose union of basins of attraction has total probability;

[21] It is worth noting that in section 31 of Landau and Lifshitz's fluid mechanics book [100], they give physical reasons why one might expect a centre bunching type of spectrum or at least a finite number of exponents near zero, in turbulent fluids.

(II) Existence of physical (SRB) measures—the attractors of the elements in D support a physical measure;

(III) Metric stability of basins of attraction—for any element in D and any of its attractors, for almost all small C^r perturbations in generic k-parameter families of dynamics, $k \in N$, there are finitely many attractors whose union of basins is nearly (Lebesgue) equal to the basin of the initial attractor; such perturbed attractors support a physical measure;

(IV) Stochastic stability of attractors—the attractors of elements of D are stochastically stable in their basins of attraction;

(V) For generic families of one-dimensional dynamics, with total probability in parameter space, the attractors are either periodic sinks or carry an absolutely continuous invariant measure.

In the portion of parameter space we examine, multiple attractors are exceedingly rare, thus yielding support for (I). This can be seen in figures 8, 9 and 12 where the initial conditions were not held fixed, yet the exponents varied continuously—leading to the conclusion that for such systems there is only one attractor. However, if the dynamics we observe are indeed turbulent-like, this is not surprising since turbulent systems rarely, if ever, exhibit multiple SRB measures. Item (II) is supported by figures 4–6 that show convergence of measure-dependent quantities. Claim (III) is particularly intriguing given the construction we utilize because of the potential of using neural networks as a probe space for the embeddings defined in section 2.1.5, thus utilizing the notions of prevalence to quantify (III) in a computational setting.

The next conjecture of Palis is directly related to the problems addressed in this work.

Conjecture 7 (Palis [9] conjecture II). *In any dimension, the diffeomorphisms exhibiting either a homoclinic tangency or a (finite) cycle of hyperbolic periodic orbits with different stable dimensions (heterodimensiopnal cycle) are C^r dense in the complement of the closure of the hyperbolic ones.*

Let us decompress this and then discuss how our results fit with it. Begin by defining the space of d-dimensional C^r diffeomorphisms as X. Next, break that space up as follows: $A = \{x \in X | x$ exhibits a homoclinic tangency or a finite cycle of hyperbolic periodic orbits with different stable dimensions $\}$ and $B = \{x \in X | x$ is hyperbolic $\}$. Thus B is the set of hyperbolic, aperiodic diffeomorphisms, and A is the set of periodic orbits or partially hyperbolic orbits. The conjecture states that A is dense in the complement of the closure of B; thus, A can be dense in B. With respect to our results, the partially hyperbolic diffeomorphisms (diffeomorphisms with homoclinic tangencies) can be dense within the set of hyperbolic diffeomorphisms. Our conjectures claim to find a subset of our one-dimensional parameter space such that partially hyperbolic diffeomorphisms will, in the limit of high dimensions, be dense. In other words, our work not only agrees with Palis's conjecture II (and subsequently his conjecture III), but our work provides evidence confirming Palis's conjectures. Of course, we do not claim to provide mathematical proofs but rather strong numerical evidence supporting Palis's ideas.

7.4. Final remarks

Finally, let us briefly summarize:

Statement of results 2 (Summary). *Assuming our particular conditions and our particular space of C^r dynamical systems as per section 1, there exists a collection of bifurcation link subsets (V) such that, in the limit of countably infinite dimensions, we have numerical evidence* *for the following.*

Conjecture 1: on the above-mentioned set V, strict hyperbolicity will be violated a-densely.

Conjecture 2: on the above-mentioned set V, the number of stable and/or unstable manifolds will change under parameter variation below numerical precision.

Conjecture 3: on the above-mentioned set V, the probability of the existence of a periodic window for a gives s on a specific parameter interval decreases inversely with dimension.

Conjecture 4: on the above-mentioned set V, hyperbolic dynamical systems are not structurally stable within numerical precision with measure one with respect to Lebesgue measure in parameter space.

In a measure-theoretic sense, hyperbolic systems occupy all the space, but the partially hyperbolic dynamical systems (with non-empty centre manifolds) can be a-dense on V. Intuitively, if there are countable dimensions—thus countable Lyapunov exponents, then one of two things can happen upon parameter variation:

(i) there would have to be a persistent homoclinic tangency—or some other sort of non-transversal intersection between stable and unstable manifolds that is persistent to parameter changes;

(ii) there can be, at most, countably many parameter points such that there are non-transversal intersections between stable and unstable manifolds.

We claim to observe (ii). We also see that for our networks, each exponent in the spectrum converges to a unique (within numerical resolution) value over a variety of initial conditions. This confirms both the usefulness and validity of our techniques and provides strong evidence for the prevalence of ergodic-like behaviour. Further, upon parameter variation, the ergodic behaviour is seemingly preserved; thus, we also have strong evidence of a prevalence of stable ergodic behaviour under parameter variation.

Acknowledgments

The authors would like to thank J R Albers, D Feldman, C McTague, J Supanich and J Thompson for a careful and thoughtful reading of this manuscript. Comments provided by K Burns were indispensable and aided in the correctness and readability of this manuscript. DJA would like to thank R A Bayliss, K Burns, W D Dechert, D Feldman, C McTague, J Robbin, C R Shalizi and J Supanich for many helpful discussions and advice. DJA would like to give special thanks to J P Crutchfield for much guidance, many fruitful discussions and support throughout this project—his support and insight contributed significantly to this paper. The computing for the project was 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. Direct support for DJA was provided by NSF grants DMR-9820816 and PHY-9910217 and DARPA Agreement F30602-00-2-0583.

References

[1] Chern S and Smale S (ed) 1970 *Global Analysis* (*Proc. Symp. Pure Math.* vol 14) (Berkeley, CA: AMS)
[2] Palis J and Smale S 1970 Structural stability theorems *Global Analysis* (*Proc. Symp. Pure Math.* vol 14) ed S Chern and S Smale (Berkeley, CA: AMS)
[3] Robbin J 1971 *Ann. Math.* **94** 447
[4] Robinson C 1971/1973 *Dynamical Systems* (Salvador: Academic) pp 443–9

[5] Mañé R 1988 *Publ. Math. IHES* **66** 161
[6] Poincaré H 1929 *Sur les propriétés des fonctions définies par les éuations aux différences partieles., Oeuvres* (Paris: Gauthier-Villars)
[7] Burns K, Pugh C, Shub M and Wilkinson A 2001 *Smooth Ergodic Theory and its Applications* (*Proc. Symp. Pure Math.* vol 69) (Berkeley, CA: American Mathematical Society) pp 327–66
[8] Pugh C and Shub M 2003 *Bull. Am. Math. Soc.* **41** 1
[9] Palis J 2000 *Asterisque* **261** 339
[10] Nitecki Z 1971 *Differentiable Dynamics: An Introduction to the Orbit Structure of Diffeomorphisms* (Boston, MA: MIT Press)
[11] Bonatti C, Diaz L and Viana M 2004 *Dynamics Beyond Uniform Hyperbolicity: A Global Geometric and Probabilistic Perspective* (Berlin: Springer)
[12] Palis J and Smale S 1970 *Global Analysis* (*Proc. Symp. Pure Math.* vol 14) (Berkeley, CA: AMS)
[13] Smale S 1967 *Bull. Am. Math. Soc.* **73** 747
[14] Newhouse S 1979 *Publ. Math. IHES* **50** 101
[15] Lorenz E N 1963 *J. Atmosph. Sci.* **20** 130
[16] Milnor J 1985 *Commun. Math. Phys.*
[17] Sommerer J and Ott E 1993 *Nature* **365** 135
[18] Kaneko K 2002 *Phys. Rev.* E **66** 055201
[19] Pugh C and Shub M 1997 *J. Complexity* **13** 125
[20] Brin M I and Pesin J B 1974 *Math. USSR Izv.* **8** 177 (English Transl.)
[21] Sompolinsky H, Crisanti A and Sommers H J 1988 *Phys. Rev. Lett.*
[22] Girko V 1994 *Random Oper. Stoch. Eqns* **2** 235
[23] Edelman A 1997 *J. Multivariate Anal.* **60** 203
[24] Bai Z D 1997 *Ann. Probab.* **25** 494
[25] Doyon B, Cessac B, Quoy M and Samuelides M 1993 *Int. J. Bifurc. Chaos* **3** 279
[26] Doyon B, Cessac B, Quoy M and Samuelides M 1994 *Acta Biotheo.* **42** 215
[27] Cessac B, Doyon B, Quoy M and Samuelides M 1994 *Physica* D **74** 24
[28] Cessac B 1995 *J. Phys. I (France)* **5** 409
[29] Moynot O and Samuelides M 2002 *Probab. Theory Relat. Fields* **123** 41
[30] Hornik K, Stinchocombe M and White H 1990 *Neural Netw.* **3** 551
[31] Sauer T, Yorke J and Casdagli M 1991 *J. Stat. Phys.* **65** 579
[32] Grassberger P 1989 *Phys. Scr.* **40** 346
[33] Kaneko K 1986 *Collapse of Tori and Genesis of Chaos in Dissipative Systems* (Singapore: World Scientific)
[34] Albers D J, Sprott J C and Crutchfield J P 2006 High-dimensional dynamics, scaling laws and general dynamics, in preparation
[35] Albers D J 2004 *PhD Thesis* University of Wisconsin in Madison
[36] Albers D J and Sprott J C 2006 Probability of a local bifurcation type from a fixed point: a random matrix perspective *J. Stat. Phys.* submitted (*Preprint* nlin.CD/0510060)
[37] Albers D J, Sprott J C and Dechert W D 1998 *Int. J. Bifurc. Chaos* **8** 1463
[38] Albers D J and Sprott J C 2006 *Physica* D submitted
[39] Takens F 1981 *Dynamical Systems and Turbulence (Warwick)* (*Lecture Notes in Mathematics* vol 898) ed D Rand and L Young (Berlin: Springer) pp 366–81
[40] Noakes L 1991 *Int. J. Bifurc. Chaos* **4** 867
[41] Gencay R and Dechert W D 1992 *Physica D* **59** 142
[42] Ott W and Yorke J 2003 *SIAM J. Appl. Dyn. Syst.* **3** 297
[43] Ott W and Yorke J A 2005 *Bull. Am. Math. Soc.* **42** 263
[44] Hunt B R, Sauer T and Yorke J A 1992 *Bull. Am. Math. Soc.* **27** 217
[45] Hunt B R, Sauer T and Yorke J A 1993 *Bull. Am. Math. Soc.* **28** 2306
[46] Albers D J, Crutchfield J P and J C Sprott 2006 Geometric variation in high-dimensional high entropy dynamical systems, in preparation
[47] Hornik K, Stinchocombe M and White H 1989 *Neural Netw.* **2** 359
[48] Adams R and Fournier J J F 2003 *Sobolev Spaces* 2nd edn (Amsterdam: Elsevier)
[49] Hebey E 1999 *Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities* (Berkeley, CA: American Mathematical Society)
[50] P L'Ecuyer 1988 *Commun. ACM* **31** 742
[51] Press W A and Teukolsky S A 1992 *Comput. Phys.* **6** 522
[52] Albers D J, Sprott J C and W D Dechert *Intelligent Engineering Systems Through Artifical Neural Networks* vol 6 (ASME) pp 17–22

[53] Amari S 2000 *Differential and Algebraic Geometry of Multilayer Perceptrons* vol 3 (*IEICE Trans. Fundamentals*)
[54] Sprott J C 2003 *Chaos and Time-series Analysis* (Oxford: Oxford University Press)
[55] Kantz H and Schreiber T 2003 *Nonlinear Time Series Analysis* 2nd edn (Cambridge: Cambridge University Press)
[56] Smale S 1966 *Am. J. Math.* **88** 491
[57] Anosov D V 1967 *Proc. Steklov Inst. Math.* **90** 1
[58] Hirsch M W Pugh C C and Shub M 1977 *Invariant Manifolds* (*Springer Lecture Notes in Mathematics*) vol 583 (Berlin: Springer)
[59] Burns K, Dolgopyat D and Pesin Y 2002 *J. Stat. Phys.* **108** 927
[60] Pugh C and Shub M 2000 *J. Eur. Math. Soc.* **2** 1
[61] Barreira L and Pesin Y 2002 *Lyapunov Exponents and Smooth Ergodic Theory* (Berkeley, CA: AMS)
[62] Katok A and Hasselblatt B 1995 *Introduction to the Modern Theory of Dynamical Systems* (*Encyclopedia of Mathematics and its Applications*) vol 54 (Cambridge: Cambridge University Press)
[63] Mané R 1987 *Ergodic Theory and Differentiable Dynamics* (Berlin: Springer)
[64] Young L-S 2002 *J. Stat. Phys.* **108** 733
[65] Albers D J, Sprott J C and Crutchfield J P 2006 Persistant chaos in high dimensions *Phys. Rev.* E at press (*Preprint* nlin.CD/0504040)
[66] Ruelle D 1979 *Publ. Math. I.H.E.S.* **50** 27
[67] Oseledec V I 1968 *Tras. Moscow Math. Soc.* **19** 197
[68] Katok A 1980 *Publ. Math. I.H.E.S.* **51** 137
[69] Ruelle D 1982 *Ann. Math.* **115** 243
[70] Pesin Y B 1976 *Soviet Math. Dokl.* **17** 196 (English Transl.)
[71] Pesin Y B 1976 *Math. USSR Izv.* **10** 1261 (English Transl.)
[72] Pesin Y B 1977 *Russ. Math. Surv.* **32** 55 (English Transl.)
[73] Robinson C 1976 *J. Diff. Eqns* **22** 28
[74] Brin M and Stuck G 2004 *Introduction to Dynamical Systems* (Cambridge: Cambridge University Press)
[75] Loéve 1977 *Probability Theory I* (Berlin: Springer)
[76] Ledrappier F and Young L-S 1985 *Ann. Math.* **122** 509
[77] Ledrappier F and Strelcyn J M 1982 *Ergod. Theory Dynam. Sys.* **2** 203
[78] Ruelle D 1976 *Am. J. Math* **98** 619
[79] Burns K, Dolgopyat D, Pesin Y and Pollicott M *Preprint*
[80] Benedicks M and Carleson L 1991 *Ann. Math* **133** 73
[81] Benedicks M and L-S Young 1993 *Invent. Math* **112** 541
[82] Sotomayor J 1973 *Dynamical Systems* pp 549–60
[83] Wiggins S 1988 *Global Bifurcations and Chaos: Analytical Methods* (*AMS*) (Berlin: Springer)
[84] Benettin G, Galgani L, Giorgilli A and J-M Strelcyn 1979 *Meccanica* **15** 9
[85] Benettin G, Galgani L, Giorgilli A and J-M Strelcyn 1979 *Meccanica* **15** 21
[86] Shimada I and Nagashima T 1979 *Prog. Theor. Phys.* **61** 1605
[87] Geist K, Parlitz U and Lauterborn W 1990 *Prog. Theor. Phys.* **83** 875
[88] Dechert W D and Gencay R 1996 *Physica D* **90** 40
[89] McDonald E J and Higham D J 2001 *Electron. Trans. Numer. Anal.* **12** 234
[90] Dieci L and Vleck E V 1995 *Appl. Numer. Math.* **17** 275
[91] Albers D J and Sprott J C contact albers@santafe.edu for notes
[92] Barreto E, Hunt B, Grebogi C and Yorke J A 1997 *Phys. Rev. Lett.* **78** 4561
[93] Banerjee S, Yorke J A and Grebogi C 1998 *Phys. Rev. Lett.* **80** 3049
[94] Potapov A and Ali M K 2000 *Phys. Lett.* A **277** 310
[95] Wheeden R L and Zygmund A 1977 *Measure and Integral Monographs and Textbooks in Pure and Applied Mathematics* vol 43 (New York: Dekker)
[96] Amari S, Nagaoka H and S-I Amari 2001 *Methods of Information Geometry (Translations of Mathematical Monographs)* (AMS)
[97] Burns K and Wilkinson A *Preprint* math.DS/0510234
[98] Wilkinson A and Dolgopyat D 2003 *Asterisque* **287** 33–60
[99] 1986 *Random Matrices and Their Applications* (*Contemoporary Mathematics*) vol 50 (AMS)
[100] Landau D and Lifshitz L 1989 *Fluid Mechanics* (Oxford: Pergamon)